

Université de Montréal

**Méthode de Simulation avec les variables
antithétiques**

par

Jean Philippe Gatarayiha

Département de mathématiques et de statistique

Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures

en vue de l'obtention du grade de

Maître ès sciences (M.Sc.)
en Statistique

juin 2007

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

**Méthode de Simulation avec les variables
antithétiques**

présenté par

Jean Philippe Gatarayiha

a été évalué par un jury composé des personnes suivantes :

Sabin Lessard

(président-rapporteur)

François Perron

(directeur de recherche)

Luc D. Adjengue

(co-directeur)

Pierre L'Ecuyer

(membre du jury)

Mémoire accepté le:

20 juin 2007

TABLE DES MATIÈRES

Liste des figures	vii
Liste des tableaux	viii
Sommaire.....	1
Summary	2
Dédicace.....	3
Remerciements	4
Introduction.....	5
Chapitre 1. Génération des variables aléatoires	9
1.1. Méthode de transformation inverse d'une distribution	9
1.1.1. Principe général	9
1.1.2. Algorithme	10
1.2. Méthode d'acceptation-rejet.....	11
1.2.1. Principe général.....	11
1.2.2. Algorithme.....	13
1.2.3. Efficacité.....	14
1.3. Génération d'une loi particulière	16
1.4. Méthode du quotient d'uniformes	19
1.4.1. Cas univarié	19
1.4.1.1. Introduction	19
1.4.1.2. Principe de la méthode.....	20

1.4.2.	Cas multivarié	29
1.4.2.1.	Méthode et principe général	29
1.4.2.2.	Algorithmes.....	32
1.5.	Méthode de Monte Carlo.....	39
1.5.1.	Introduction	39
1.5.2.	Efficacité.....	40
1.5.3.	Intégration de Monte Carlo.....	40
1.5.3.1.	Méthode de Monte Carlo de « Noir ou Blanc »	40
1.5.3.2.	Algorithme	44
1.5.3.3.	Méthode de Monte Carlo moyenne échantillonnale.....	44
1.5.3.4.	Algorithme	45
1.5.3.5.	Efficacité de la méthode de Monte Carlo	47
Chapitre 2.	Techniques de réduction de la variance	49
2.1.	Introduction	49
2.2.	Échantillonnage stratifié.....	49
2.2.1.	Introduction et principe général.....	49
2.3.	Échantillonnage pondéré	56
2.3.1.	Principe général	56
2.3.2.	Choix optimal de la fonction d'importance	58
2.3.3.	Algorithme	59
2.4.	Variables de contrôle.....	61
2.4.1.	Idée générale.....	61
2.5.	Variables antithétiques.....	64
2.5.1.	Idée générale.....	64
2.5.2.	Variables antithétiques pour des fonctions monotones	66

2.5.2.1.	Les fonctions g et h sont monotones dans chaque composante de x et y , $x, y \in \mathbb{R}^n$	72
2.5.2.2.	Les fonctions g et h sont monotones et croissantes pour certaines composantes et monotones décroissantes pour d'autres composantes	73
2.5.2.3.	Les fonctions g et h ne sont pas monotones	76
2.5.3.	Variables antithétiques pour des fonctions non monotones	79
2.5.3.1.	Introduction	79
2.5.3.2.	La version modifiée du théorème de Handscomb	81
2.5.3.3.	Procédure optimale pour $n = 2$	83

Chapitre 3. Nouvelle méthode de simulation avec variables antithétiques

89

3.1.	Fonctions monotones	89
3.1.1.	Principe général et exemple	89
3.1.1.1.	Cas où $m = 1$	89
3.1.1.2.	Généralisation : cas où $m > 1$	91
3.1.2.	Formules des variances pour la fonction identité et exponentielle	93
3.1.3.	Résultats théoriques	96
3.1.4.	Applications de la méthode	99
3.1.4.1.	Fonction identité	100
3.1.4.2.	Modèle exponentiel	103
3.1.5.	Résultats de simulation	122
3.1.5.1.	Modèle exponentiel	122
3.1.5.2.	Espérance de X pour le modèle uniforme	124
3.1.5.3.	Espérance de e^X pour le modèle uniforme	125
3.1.5.4.	Espérance de $\cos(tX)$ pour le modèle uniforme	126
3.1.6.	Discussions	127
3.1.6.1.	Cas où X n'est pas de loi uniforme sur $(0, 1)$	127

3.1.6.2. Cas où la fonction f n'est pas croissante	128
Conclusion	129
Annexe	132
- Estimation de $E(X)$ pour le modèle exponentiel	132
- Estimation de $E(X)$ par le modèle uniforme.....	132
- Estimation de $E(e^X)$ pour le modèle uniforme.....	132
- Estimation de $E(\cos(tX))$ pour le modèle uniforme, $t = 20$	133
Programme Splus	137
Bibliographie	140

LISTE DES FIGURES

1.1	Méthode acceptation-rejet de J. von Neumann	12
1.2	Région d'acceptation \mathcal{D} d'une normale obtenue par la méthode du quotient d'uniformes.....	28
1.3	Graphique : Méthode de Noir ou Blanc où $a = -1,7$ et $b = 0$	43
3.1	Variances : $E(X)$ par le modèle exponentiel pour $n = 200$	123
3.2	Variances : $E(X)$ pour le modèle uniforme pour $n = 200$	124
3.3	Variances : $E(e^X)$ pour le modèle uniforme avec $n = 200$	125
3.4	Variances : $E(\cos(tX))$ pour le modèle uniforme avec $n = 200, t = 20$.	126

LISTE DES TABLEAUX

2.1	Tableau des probabilités conjointes de (X, Y)	56
2.2	Tableau des espérances et variances de $(Y X = i)$	56
3.1	Tableau 1 des formules des variances : cas stratifié, $m = 1$	93
3.2	Tableau 2 : Résumé des formules des variances : Cas Stratifié pour $f(x) = x$ et $f(x) = e^{tx}$ ($m = 1$)	95
3.3	Annexe 1 : Tableau des variances et des moyennes pour $N = 3000$	133
3.4	Annexe 2 : Tableau des variances et des moyennes pour $N = 50000$..	134
3.5	Annexe 3 : Tableau des variances et des moyennes ($E(X)$) pour le modèle uniforme avec $N = 1000$	134
3.6	Annexe 4 : Tableau des variances et des moyennes de ($E(\exp(X))$) par le modèle uniforme avec $N = 1000$	135
3.7	Annexe 5 : Tableau des variances et des moyennes de ($E(\cos(tX))$) pour le modèle uniforme avec $N = 1000$ et $t = 20$	136

SOMMAIRE

Dans ce mémoire, nous travaillons sur une méthode de simulation de Monte Carlo qui utilise des variables antithétiques pour estimer une intégrale de la forme :

$$I = \int_0^1 f(x)dx$$

où f peut être une fonction monotone, non monotone ou une fonction difficile à simuler.

L'idée principale de la méthode qu'on propose est de subdiviser l'intervalle $[0,1]$ en m sections dont chacune est subdivisée en l sous intervalles. Cette technique se fait en plusieurs étapes et à chaque fois qu'on passe à l'étape supérieure la variance diminue. C'est-à-dire que la variance obtenue à la k^e étape est plus petite que celle trouvée à la $(k-1)^e$ étape ce qui nous permet également de rendre plus petite l'erreur d'estimation car l'estimateur de I est sans biais. L'objectif est de trouver m , le nombre optimal de sections, qui permet de trouver cette diminution de la variance.

Mots clés : Estimateur sans biais, génération de variables aléatoires, Simulation de Monte Carlo, Variables antithétiques, Réduction de la variance.

SUMMARY

In this master thesis, we consider simulation methods based on antithetic variates for estimate integrals of the form :

$$I = \int_0^1 f(x)dx$$

where f is a monotonic function, not a monotonic function or a function difficult to integrate.

The main idea consists in subdividing the interval $(0, 1]$ in m sections of which each one is subdivided in l subintervals. This method is done recursively. At each step the variance decreases, i.e. the variance obtained at the k^{th} step is smaller than that is found at the $(k - 1)^{th}$ step. This allows us to reduce the error in the estimation because the estimator of I is unbiased. The objective is to optimize m .

Key Words : Unbiased estimator, random numbers, Monte Carlo simulation, antithetic variates, variance reduction.

DÉDICACE

- A ma très chère épouse Alice KAYIHURA GATARAYIHA,
- A ma famille,
- Et à ma belle famille.

REMERCIEMENTS

C'est tout à fait logique que je commence par remercier mon Directeur de recherche François Perron pour l'énorme investissement de temps et d'effort qu'il a su faire. Côtoyer quelqu'un d'aussi ouvert socialement et généreux intellectuellement m'a beaucoup aidé non seulement à finir ce travail de recherche mais aussi m'a permis d'aller jusqu'au bout dans mes études de maîtrise.

Dans la même ordre d'idée je remercie mon Co-Directeur Luc Adjengué, Professeur à l'Ecole Polytechnique de Montréal pour son aide tant intellectuelle que financière.

Je profite de cette belle occasion pour remercier toute personne qui m'a aidé à intégrer la société canadienne en général et la société universitaire québécoise en particulier, entre autres tous mes Professeurs de statistique au département de mathématiques et de statistique qui m'ont accepté comme étudiant et tous les étudiants de maîtrise qui m'ont toujours considéré comme collègue. Mes remerciements les plus sincères au Professeur Jean-François Angers et à Miguel Chagnon qui m'ont donné l'opportunité de présenter un projet au congrès de la SSC du mois de juin 2004.

Je tiens à remercier au fond de mon coeur mon épouse Alice Kayihura d'avoir pu supporter cette si longue séparation et surtout de m'avoir toujours encouragé.

Un grand merci à Danny Nzaramba de m'avoir aidé et encouragé à aller poursuivre mes études. J'en suis reconnaissant.

Cela serait injuste d'oublier de remercier mes amis de longue date Jean Marie Vianney Gatete de m'avoir trouvé une place dans son entreprise et Roger Remera de m'avoir offert en cadeau un livre de statistique et probabilité avant même que je me décide à faire la maîtrise en statistique.

INTRODUCTION

Le nom et le développement systématique des méthodes de simulation dites de Monte Carlo datent de 1944 environ et est lié au projet de la première bombe atomique. À l'aide de simulations directes on traitait des problèmes probabilistes reliés à la diffusion aléatoire des neutrons dans les matériaux fissiles. A ce stade primitif de ces recherches, von Neumann et Ulam ont amélioré ces simulations directes au moyen de certaines techniques de réduction de la variance. Il a fallu attendre Harris et Kahn (1948) pour un développement systématique de ces idées (Monte carlo method, 1964)

Plusieurs calculs de Monte Carlo qui conduisent à des résultats quantitatifs peuvent être considérés comme un problème d'estimation d'une intégrale multiple. Le problème de l'évaluation des intégrales fournit effectivement une plateforme utile pour mettre en relief différentes techniques de Monte Carlo d'application générale. On prendra au départ pour plus de simplicité l'intégrale à une dimension suivante :

$$I = \int h(x)f(x)dx \quad \text{où } f \text{ est une fonction de densité.}$$

Les différentes approches qui reposent sur la loi des grands nombres pour évaluer cette intégrale I consistent à évaluer h en des points $x_i, i = 1, 2, \dots, n$ où le choix des x_i est déterminé par la réalisation d'une variable aléatoire X de densité f .

C'est-à-dire, si

$$I = E_f(h(X)) = \int h(x)f(x)dx \tag{0.0.1}$$

alors

$$\hat{I} = \frac{1}{m} \sum_{j=1}^m h(x_j). \quad (0.0.2)$$

La convergence de \hat{I} vers I lorsque m devient grand vient de la loi forte des grands nombres ainsi l'expression (0.0.2) converge presque sûrement vers $E_f(h(x))$.

Si h^2 est intégrable par rapport à f alors

$$\text{Var}(\hat{I}) = \frac{1}{m} \left[\int_{\mathcal{X}} h^2(x) f(x) dx - (E_f(h(X)))^2 \right] \quad (0.0.3)$$

Dans cette approche, il faut être capable d'obtenir des réalisations indépendantes de X selon la densité f et cela devient parfois trop difficile. Dans le cas où on ne parvient pas à générer selon f on peut introduire une autre méthode dite d'échantillonnage pondéré (*Importance sampling*). Cette méthode nécessite l'ajout d'une autre fonction de densité dite densité instrumentale et fait intervenir des poids liés au rapport entre la densité f et la densité instrumentale et leur propriété sera expliquée dans le troisième chapitre de ce document.

Comme on l'a déjà mentionné dans le premier paragraphe, l'amélioration récente des techniques de réduction de la variance a rendu les méthodes de Monte Carlo plus efficaces en divers domaines de la statistique. Parmi les différentes techniques, on va revenir sur la technique d'échantillonnage pondéré et on va expliquer la technique d'échantillonnage stratifié qui consiste à partitionner l'espace en plusieurs sous-ensembles appelés strates. On effectue ensuite les tirages de Monte Carlo élémentaires séparément dans chaque strate.

Si on part d'un estimateur sans biais on peut chercher à minimiser la variance. Pour faire face à ce problème d'optimalité, on va aborder deux autres techniques de réduction de la variance qui introduisent une structure de corrélation sur les variables qui sont générées dans l'échantillon de Monte Carlo. On s'intéressera plus particulièrement aux variables de contrôle et aux variables antithétiques.

Cependant, la comparaison de ces méthodes de simulation ne peut pas se faire uniquement au travers des variances des estimateurs résultants car il faut également prendre en compte les temps de calcul et la complexité de l'algorithme

d'où l'utilité des approches de Monte Carlo par chaînes de Markov (MCMC) qui permettent de gagner du temps.

Le but de ce mémoire est d'introduire plusieurs techniques de simulation venant de diverses sources variées, d'expliquer ces techniques et de les appliquer sur des exemples. Ce mémoire sera constitué de trois grandes parties :

- La première partie sera consacrée aux trois méthodes de génération de variables aléatoires suivantes : méthode de transformation inverse, méthode d'acceptation rejet et méthode du quotient d'uniformes. Ensuite on va expliciter les algorithmes de ces méthodes. On va donner aussi des variantes où il faut modéliser les algorithmes sur des cas particuliers comme on en trouve dans Devroye (1986).

- La deuxième partie mettra en évidence les quatre techniques de réduction de la variance suivantes : échantillonnage stratifié, échantillonnage pondéré, variables de contrôle et variables antithétiques.

La plus grande partie de ce document sera consacrée aux variables antithétiques, dont le nom sinon l'idée a été introduit par Hammersley et Mauldon (1956). Au fil des années d'autres auteurs ont continué à approfondir ou améliorer cette technique. Hammersley et Morton (1956), ont proposé le premier résultat fondamental sur les variables antithétiques et qui a été démontré par Handscomb (1958) d'où le nom du théorème de Handscomb. En 1957, Morton donne la généralisation de la variable antithétique dans l'évaluation de l'intégrale. Wilson (1979) a pu généraliser et démontrer le théorème de Hammersley dans le cas des fonctions non bornées et pour $n \geq 2$. Fishman, Georges et Huang (1982) proposent les variables antithétiques révisées avec l'introduction de l'effet de rotation et de réflexion de l'échantillon, ce qui leur a permis de trouver la variance minimale. Rubinstein et Samorodnitsky (1986), proposent la version modifiée du théorème de variables antithétiques de Handscomb. Ils ont pu démontrer l'existence d'un minimum exact contrairement à l'infimum atteint par Hammersley, Morton et Handscomb.

- Dans la troisième partie, nous allons proposer une autre méthode de simulation avec variables antithétiques sur les fonctions monotones, non monotones

et sur des fonctions difficiles à simuler. La méthode de simulation que nous introduisons est appelée latin hypercube sampling en anglais. Il est très difficile de trouver des démonstrations où on donne une forme explicite du gain relatif réalisé par cette méthode par rapport à un échantillonnage simple. Dans ce mémoire nous donnons une démonstration originale où le gain prend une forme explicite. De plus, nous dérivons des expressions analytiques pour ce gain dans des situations où il est possible d'obtenir des expressions explicites. Nous faisons aussi des études numériques du gain à l'aide de simulations pour des situations où nous ne parvenons pas à obtenir une réponse de façon analytique.

Chapitre 1

GÉNÉRATION DES VARIABLES ALÉATOIRES

Ce chapitre comporte deux parties essentielles : la première partie traite de quelques procédures pour générer des variables selon différentes distributions : méthode de transformation inverse, méthode d'acceptation-rejet, méthode du quotient d'uniformes et la génération de lois particulières.

La seconde partie de ce chapitre décrit en détail la méthode de Monte Carlo noir ou blanc et la méthode échantillonnale de Monte Carlo.

1.1. MÉTHODE DE TRANSFORMATION INVERSE D'UNE DISTRIBUTION

1.1.1. Principe général

Cette méthode est la plus facile et la plus directe pour générer les variables aléatoires. Elle a l'inconvénient d'exiger la connaissance de F^{-1} où F est la fonction de répartition.

Théorème 1.1.1. *Soit F une fonction de répartition définie sur un intervalle $[a, b]$ avec fonction inverse (généralisée) définie par :*

$$F^{-1}(u) = \inf\{x \in [a, b] : F(x) \geq u\} \text{ où } 0 \leq u \leq 1.$$

La variable $F^{-1}(U)$ est une variable aléatoire de fonction de répartition F lorsque U est de loi uniforme sur $(0, 1)$.

Démonstration

L'inverse généralisé vérifie, $F(F^{-1}(u)) \geq u$ et $F^{-1}(F(x)) \leq x$, pour tout $u \in [0, 1]$, $x \in F^{-1}([0, 1])$.

Par conséquent $\{(u, x) : F^{-1}(u) \leq x\} = \{(u, x) : u \leq F(x)\}$,
d'où $F(x) = P[F^{-1}(u) \leq x]$, car $P[F^{-1}(u) \leq x] = P[u \leq F(x)]$ en terme d'événement.

1.1.2. Algorithme

L'algorithme pour générer une observation d'une variable aléatoire X par l'inverse généralisé va comme suit :

- (1) Générer aléatoirement u , une réalisation d'une loi uniforme sur $(0, 1)$.
- (2) Poser $x = F^{-1}(u)$.

Exemple 1.1.1.

Si X est de loi exponentielle(λ)

alors $F(x) = 1 - \exp(-\lambda x)$, $x > 0$,

$$\text{d'où } F^{-1}(u) = \frac{-\log(1-u)}{\lambda}.$$

Algorithme

- (1) Générer aléatoirement u suivant une loi uniforme sur $(0, 1)$.
- (2) Poser $x = -\log(1-u)/\lambda$.

Exemple 1.1.2.

Soit une distribution de Cauchy de paramètre (α, β) notée $C(\alpha, \beta)$ dont la fonction de densité est :

$$f(x) = \frac{\beta}{\pi [\beta^2 + (x - \alpha)^2]}$$

avec $\alpha, x \in \mathbb{R}, \beta > 0$.

La fonction de répartition est :

$$F(x) = \frac{1}{2} + \pi^{-1} \tan^{-1} \left(\frac{x - \alpha}{\beta} \right).$$

En appliquant la méthode de transformation inverse nous obtenons :

$$\begin{aligned} X &= F_X^{-1}(U), \\ &= \alpha + \beta \tan \left[\pi \left(U - \frac{1}{2} \right) \right], \\ &= \alpha - \frac{\beta}{\tan(\pi U)}. \end{aligned}$$

Algorithme

- (1) Générer u selon une loi uniforme sur $(0, 1)$.
- (2) Poser $x = \alpha - \beta / \tan(\pi u)$.

1.2. MÉTHODE D'ACCEPTATION-REJET

1.2.1. Principe général

Cette méthode a été élaborée pour la première fois par John von Neumann(1951). La méthode de rejet peut utiliser la génération de n'importe quel type de variable aléatoire. Elle consiste à générer des observations suivant une distribution de fonction de densité instrumentale et à ensuite éliminer une certaine proportion de ces observations à partir d'un test de manière à se ramener à des observations qui suivent la distribution désirée.

Plus précisément, soient f la densité désirée de la variable aléatoire à générer, g une fonction de densité instrumentale et C une constante positive telle que $f(x) \leq Cg(x)$ pour toute valeur de x .

Pour effectuer la méthode d'acceptation-rejet, on considère X une variable générée selon une loi de densité f telle que :

$$f(x) = Ch(x)g(x),$$

où $C \geq 1$, g est une fonction densité, $\phi(x) = Cg(x)$ est une fonction qui majore la densité f (c.-à-d. $f(x) \leq Cg(x), \forall x$), et $0 < h(x) = f(x)/\phi(x) \leq 1$ (fig.(1.1)).

Alors on va générer deux variables U et Y selon une loi $\mathcal{U}(0, 1)$ et une loi de densité g respectivement et on va tester si $U \leq h(Y)$.

- (1) Si l'inégalité est vérifiée alors on accepte que Y est généré selon une loi de densité f .
- (2) Sinon on oublie et on recommence de nouveau.

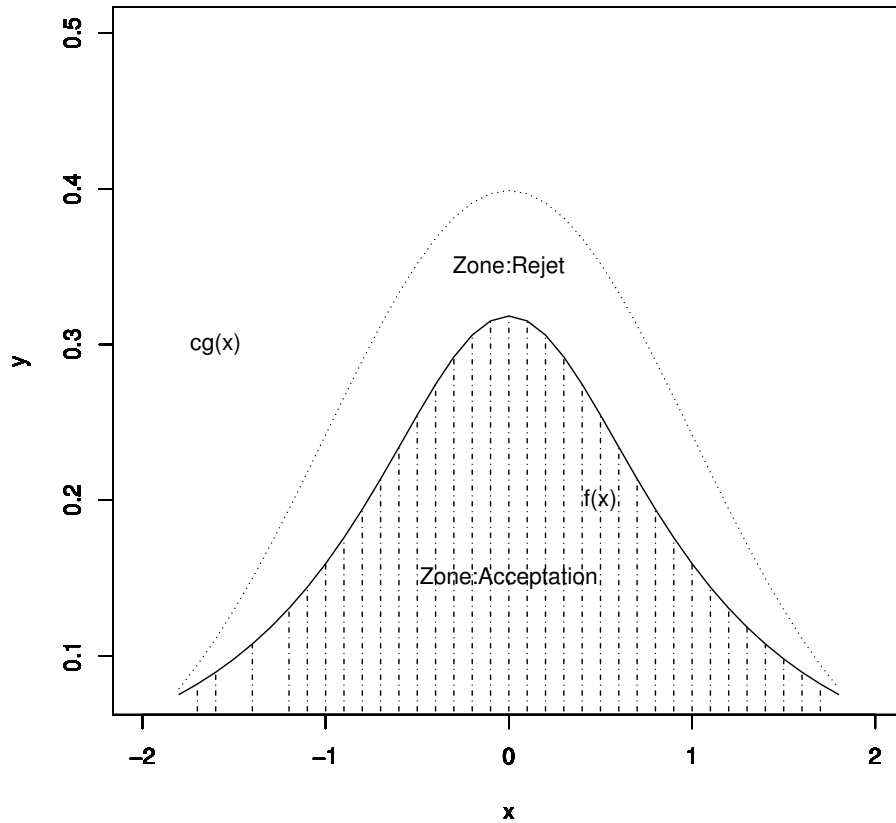


FIGURE 1.1. Méthode acceptation-rejet de J. von Neumann

La théorie de base de la méthode d'acceptation-rejet provient du théorème suivant :

Théorème 1.2.1. *Soit X une variable aléatoire générée selon une loi de densité f telle que*

$$f(x) = Ch(x)g(x)$$

où $C \geq 1$, g est aussi une densité, et $0 < h(x) \leq 1$ pour tout x . Soient U et Y deux variables aléatoires distribuées selon une loi uniforme sur $(0, 1)$ et une loi de densité g respectivement. Nous obtenons que la loi conditionnelle d'une observation sachant que le test a été accepté a comme densité f .

Démonstration

Pour tout $x \in \mathbb{R}$ nous avons

$$\begin{aligned} P[Y \leq x | U \leq h(Y)] &= \frac{\int_{-\infty, x] \int_{[0, h(y)]} g(y) du dy}{\int_{-\infty, +\infty) \int_{[0, h(y)]} g(y) du dy} \\ &= \frac{\int_{-\infty, x] h(y) g(y) dy}{\int_{-\infty, +\infty) h(y) g(y) dy} \\ &= \int_{-\infty, x] f(y) dy. \end{aligned}$$

1.2.2. Algorithme

L'algorithme sera donc :

- (1) Générer U selon une loi uniforme sur $(0, 1)$.
- (2) Générer Y selon une loi de densité g .
- (3) Si $U \leq h(Y)$ alors Y est généré selon une loi de densité f .
- (4) Sinon on oublie et c'est le retour à l'étape 1.

L'algorithme se termine lorsqu'on a accepté n fois pour un échantillon de taille n .

1.2.3. Efficacité

L'efficacité de cette méthode est définie comme la moyenne de la fréquence relative des points acceptés et se mesure par le rapport entre les surfaces sous les courbes f et Cg . Comme f et g sont des fonctions densité, ce rapport vaut $1/C$.

Exemple 1.2.1. *Fishman (1976)*

Générer une variable aléatoire selon la loi $\text{Gamma}(\alpha, 1)$ avec $\alpha \geq 1$. On peut prendre la densité d'une exponentielle comme densité instrumentale, d'où

$$g(x) = \frac{1}{\alpha} \exp(-x/\alpha), \quad g \text{ est une densité.}$$

Ainsi,

$$h(x) = x^{\alpha-1} \frac{\exp[-x(1-1/\alpha)]}{\alpha^{\alpha-1} \exp(1-\alpha)},$$

et

$$C_0 = \frac{\alpha^\alpha \exp(1-\alpha)}{\Gamma(\alpha)}, \alpha \geq 1,$$

devient la plus petite valeur qu'on puisse donner à C pour que l'algorithme fonctionne.

Il est évident que la condition $U \leq h(Y)$, où la variable aléatoire Y provient d'une loi exponentielle($1/\alpha$), peut s'écrire comme $V_2 \geq (\alpha - 1)(V_1 - \ln(V_1) - 1)$ où V_1 et V_2 sont des variables aléatoires d'une loi exponentielle(1). L'algorithme sera donc :

- (1) Générer V_1 et V_2 selon une loi exponentielle(1).
- (2) Si $V_2 < (\alpha - 1)(V_1 - \ln(V_1) - 1)$, retourner à l'étape précédente.
- (3) Sinon V_1 est la valeur retenue comme variable de loi $\Gamma(\alpha, 1)$.

Exemple 1.2.2. *Soit*

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$

la densité d'une normale centrée réduite. Cette densité est symétrique par rapport à l'origine.

Considérons la densité de la loi de Laplace comme densité instrumentale. Par la méthode d'acceptation-rejet on a :

$$f(x) = Ch(x)g(x),$$

$$\text{où } g(x) = \frac{1}{2} \exp(-|x|),$$

$$C_0 = \sqrt{\frac{2e}{\pi}},$$

$$h(x) = \exp \left[-\frac{(|x| - 1)^2}{2} \right].$$

L'efficacité de la méthode est égale à $\sqrt{\pi/2e} = 0,76$.

La condition d'acceptation

$$U \leq h(Y) \quad \text{devient} \quad U \leq \exp[-(Y - 1)^2/2],$$

elle équivaut à

$$-\ln U > (|Y| - 1)^2/2$$

où Y est de loi de Laplace.

1.3. GÉNÉRATION D'UNE LOI PARTICULIÈRE

Certaines fonctions de répartition ne sont pas explicites, on peut utiliser des approximations de ces fonctions, mais on risque d'accumuler des erreurs.

Dans cette section on donne des algorithmes pour travailler avec ces lois particulières. On va considérer deux exemples : la loi bêta avec l'approche de Jöhnk (1976), et la loi normale par la procédure classique de Box et Muller (1958).

L'approche de Jöhnk pour la loi bêta

Cette procédure est basée sur le théorème suivant :

Théorème 1.3.1. *Soient U_1 et U_2 deux variables aléatoires indépendantes et uniformes sur $(0, 1)$ et soit $Y_1 = U_1^{1/\alpha}$ et $Y_2 = U_2^{1/\beta}$. Si $Y_1 + Y_2 \leq 1$, alors*

$$X = \frac{Y_1}{Y_1 + Y_2}$$

est de loi bêta de paramètres (α, β) .

Démonstration

Nous avons :

$$f_{Y_1}(y_1) = \alpha y_1^{\alpha-1}, 0 \leq y_1 \leq 1,$$

$$f_{Y_2}(y_2) = \beta y_2^{\beta-1}, 0 \leq y_2 \leq 1,$$

et

$$f_{Y_1, Y_2}(y_1, y_2) = \alpha \beta y_1^{\alpha-1} y_2^{\beta-1}.$$

Soit $X = Y_1/W$ avec $W = Y_1 + Y_2$. Le Jacobien de transformation devient

$$J = \begin{vmatrix} \frac{\partial y_1}{\partial x} & \frac{\partial y_1}{\partial w} \\ \frac{\partial y_2}{\partial x} & \frac{\partial y_2}{\partial w} \end{vmatrix} = \begin{vmatrix} w & x \\ -w & 1-x \end{vmatrix} = w$$

$$\text{et } f_{X,W}(x, w) = \alpha \beta x^{\alpha-1} (1-x)^{\beta-1} w^{\alpha+\beta-1},$$

$$0 \leq x \leq 1, 0 \leq w \leq 2.$$

Lorsqu'on conditionne sur l'événement $0 \leq W \leq 1$ la densité conditionnelle devient

$$f_X(x|0 \leq W \leq 1) = \frac{f_{X,W}(x, 0 \leq W \leq 1)}{P(0 \leq W \leq 1)}, \quad (1.3.1)$$

$$\text{et } f_{X,W}(x, 0 \leq W \leq 1) = \int_0^1 f_{X,W}(x, w) dw \quad (1.3.2)$$

$$= \frac{\alpha\beta}{\alpha + \beta} x^{\alpha-1} (1-x)^{\beta-1}, \quad (1.3.3)$$

$$\text{avec } 0 \leq x \leq 1.$$

$$\begin{aligned} P(0 \leq W \leq 1) &= \int_0^1 f_{X,W}(x, 0 \leq W \leq 1) dx \\ &= \frac{\alpha\beta}{\alpha + \beta} \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}. \end{aligned} \quad (1.3.4)$$

En substituant (1.3.4) et (1.3.3) dans (1.3.1), nous obtenons

$$f_X(x|0 \leq W \leq 1) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}.$$

L'efficacité de la méthode est égale à

$$\frac{1}{C} = P(Y_1 + Y_2 \leq 1) = P(0 \leq w \leq 1) = \frac{\alpha\beta}{\alpha + \beta} \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}.$$

Pour des grandes valeurs de α et β la procédure de Jöhnk n'est pas très efficace.

L'approche de Box et Muller pour la loi normale

Soient U_1 et U_2 deux variables aléatoires indépendantes générées selon une loi uniforme sur $(0, 1)$. Les variables

$$Z_1 = (-2 \ln(U_1))^{1/2} \cos 2\pi U_2$$

$$Z_2 = (-2 \ln(U_1))^{1/2} \sin 2\pi U_2$$

sont aussi indépendantes et sont générées selon la normale centrée réduite. On peut réécrire les deux variables de la façon suivante :

$$Z_1 = (2V)^{1/2} \cos 2\pi U$$

$$Z_2 = (2V)^{1/2} \sin 2\pi U$$

où V suit la loi exponentielle de paramètre 1 et $U_2 = U$.

$$\text{Donc on a } Z_1^2 + Z_2^2 = 2V \quad \text{et} \quad \frac{Z_1}{Z_2} = \tan(2\pi U)$$

Le Jacobien de la transformation est donné par

$$\begin{aligned} J &= \begin{vmatrix} \frac{\partial u}{\partial z_1} & \frac{\partial u}{\partial z_2} \\ \frac{\partial v}{\partial z_1} & \frac{\partial v}{\partial z_2} \end{vmatrix} = \begin{vmatrix} -\frac{z_2(\cos 2\pi U)^2}{2\pi z_1^2} & \frac{(\cos 2\pi U)^2}{2\pi z_1} \\ z_1 & z_2 \end{vmatrix} \\ &= \begin{vmatrix} -\frac{z_2}{4\pi v} & \frac{z_1}{4\pi v} \\ z_1 & z_2 \end{vmatrix} = -\frac{1}{4\pi v} [z_2^2 + z_1^2] \\ &= -\frac{2v}{4\pi v} = -\frac{1}{2\pi}. \end{aligned}$$

Ainsi,

$$\begin{aligned}
 f_{Z_1, Z_2}(z_1, z_2) &= f_{Z, V}(z, v) |J| \\
 &= \frac{1}{2\pi} \exp\left(-\frac{z_1^2 + z_2^2}{2}\right) \\
 &= \frac{1}{\sqrt{(2\pi)}} \exp\left(-\frac{z_1^2}{2}\right) \frac{1}{\sqrt{(2\pi)}} \exp\left(-\frac{z_2^2}{2}\right) \\
 &= f_{Z_1}(z_1) f_{Z_2}(z_2).
 \end{aligned}$$

Finalement, Z_i , $i = 1, 2$, sont des variables de lois normales centrées de variance 1 et les variables Z_i , $i = 1, 2$, sont indépendantes.

Algorithme

- (1) Générer U_1 et U_2 selon des lois uniformes indépendantes sur $(0, 1)$.
- (2) Calculer Z_1 et Z_2 en utilisant les formules suivantes :

$$\begin{aligned}
 Z_1 &= (-2 \ln(U_1))^{1/2} \cos 2\pi U_2 \\
 Z_2 &= (-2 \ln(U_1))^{1/2} \sin 2\pi U_2.
 \end{aligned}$$

1.4. MÉTHODE DU QUOTIENT D'UNIFORMES

1.4.1. Cas univarié

1.4.1.1. Introduction

La méthode du quotient d'uniformes a été proposée pour la première fois par Kinderman et Monahan (1977) et plus tard en 1979 ils ont proposé l'application de cette méthode à la distribution de Student puis Cheng et Feast (1979) l'ont appliqué à la distribution Gamma.

Stefănescu et Văduva (1987) ont généralisé cette méthode à des distributions multivariées.

L'efficacité de cette méthode résulte du fait de son algorithme qui est court, facile à comprendre et surtout très rapide au moment de son exécution.

1.4.1.2. Principe de la méthode

La méthode de quotient d'uniformes consiste à générer un point uniformément dans une certaine région \mathcal{D} incluse dans un ensemble plus simple tel que le rectangle, dans lequel il est facile de générer ce vecteur. La variable générée avec la distribution désirée est obtenue en faisant le rapport de deux coordonnées de ce point. Le théorème suivant Fishman (1996) est le cas général du théorème de Kinderman et Monahan.

Théorème 1.4.1. *Soit r une fonction sommable sur \mathbb{R} à valeur dans $[0, \infty)$. Considérons l'ensemble \mathcal{D} donné par*

$$\mathcal{D} = \left\{ (x, y) : 0 \leq x \leq \sqrt{r(a_1 + a_2(y/x))} \right\}, \quad -\infty < a_1 < \infty, a_2 > 0.$$

Soit (X, Y) un vecteur uniformément distribué sur $\mathcal{B} \supseteq \mathcal{D}$. Posons $W = X^2$ et $Z = a_1 + a_2 Y/X$. À la distribution conditionnelle de Z étant donné l'événement $W \leq r(X)$, correspond une densité f_Z et f_Z est proportionnel à r , c'est-à-dire,

$$f_Z(z) = r(z)/d \quad \text{avec} \quad -\infty < z < \infty,$$

où

$$d = \int_{-\infty}^{\infty} r(z) dz = 2a_2 m(\mathcal{D})$$

avec $m(\mathcal{D})$ qui est la mesure de \mathcal{D} . En pratique on choisira \mathcal{B} le plus petit rectangle qui contient \mathcal{D} .

Démonstration

Si $(X, Y) \in \mathcal{D}$ alors les variables X et Y ont la densité jointe

$$f_{X,Y}(x, y) = 1/|\mathcal{D}| \quad \text{pour} \quad (x, y) \in \mathcal{D}$$

et

$$f_{W,Z}(w, z) = J f_{X,Y}(\sqrt{w}, (z - a_1)\sqrt{w}/a_2), \quad 0 \leq w \leq r(z), \quad -\infty < z < \infty,$$

où J est le jacobien,

$$J = \left| \begin{array}{cc} \frac{\partial x}{\partial w} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial w} & \frac{\partial y}{\partial z} \end{array} \right| = \left| \begin{array}{cc} \frac{1}{2\sqrt{w}} & 0 \\ \frac{z-a_1}{2a_2\sqrt{w}} & \frac{\sqrt{w}}{a_2} \end{array} \right| = \frac{1}{2a_2},$$

de sorte que

$$f_Z(z) = \int_0^{r(z)} f_{W,Z}(w, z) dw = \frac{r(z)}{2a_2 m(\mathcal{D})}, \quad -\infty < z < \infty$$

et

$$\int_{-\infty}^{\infty} f_Z(z) dz = \frac{1}{2a_2 m(\mathcal{D})} \int_{-\infty}^{\infty} r(z) dz = 1.$$

Les équations

$$x(z) = r^{1/2}(z)$$

et

$$y(z) = (z - a_1)x(z)/a_2 = (z - a_1)r^{1/2}(z)/a_2$$

définissent la frontière de \mathcal{D} , et chaque point de \mathcal{D} est dans le rectangle

$$\mathcal{B} = \{(x, y) : 0 \leq x \leq x^*, \quad y_* \leq y \leq y^*\},$$

où

$$x^* = \sup_z r^{1/2}(z),$$

et

$$y_* = \inf_z \left[(z - a_1)r^{1/2}(z)/a_2 \right] \quad \text{et} \quad y^* = \sup \left[(z - a_1)r^{1/2}(z)/a_2 \right].$$

Algorithme

- (1) Choisir $r \propto f$, $a_1 \in (-\infty, \infty)$ et $a_2 \in (0, \infty)$.
- (2) Calculer $x^* \leftarrow \sup_z r^{1/2}(z)$, $y_* \leftarrow \inf_z \left[(z - a_1)r^{1/2}(z)/a_2 \right]$,
 $y^* \leftarrow \sup \left[(z - a_1)r^{1/2}(z)/a_2 \right]$.
- (3) Générer aléatoirement X selon une loi $\mathcal{U}(0, x^*)$.
- (4) Générer aléatoirement Y selon une loi $\mathcal{U}(y_*, y^*)$.
- (5) Poser $Z \leftarrow a_1 + a_2 Y/X$.
- (6) Si $X^2 \leq r(Z)$ on accepte Z , sinon on retourne à l'étape (3).

Donc si $X^2 \leq r(a_1 + a_2 Y/X)$, alors $(X, Y) \in \mathcal{D}$, et $Z = a_1 + a_2 Y/X$ a une densité f_Z désirée.

Cependant la probabilité d'accepter Z est donnée par

$$\begin{aligned} P[X^2 \leq r(a_1 + a_2 Y/X)] &= P[(X, Y) \in \mathcal{D}] = m(\mathcal{D})/m(\mathcal{B}) \\ &= \frac{2a_2 x^* (y^* - y_*)}{d}. \end{aligned}$$

Remarque

La région \mathcal{B} peut prendre une forme autre que rectangulaire afin de rendre le rapport $|\mathcal{D}|/|\mathcal{B}|$ plus près de l'unité, on doit équilibrer cette réduction en tenant compte du coût pour générer X et Y sur une région plus compliquée.

On remarque que les relations $x^* \leftarrow \sup_z r^{1/2}(z)$, $y_* \leftarrow \inf_z [(z - a_1)r^{1/2}(z)/a_2]$, et $y^* \leftarrow \sup [(z - a_1)r^{1/2}(z)/a_2]$ dépendent des a_i , $i=1,2$, donc les a_i permettent de trouver la frontière du rectangle \mathcal{B} qui contient la région \mathcal{D} .

Le théorème original suivant proposé par Kinderman et Mohanan (1977) est un cas particulier du théorème (1.4.1) avec $a_1 = 0$ et $a_2 = 1$.

Théorème 1.4.2. (*Kinderman et Mohanan*)

$$\text{Soit } \mathcal{D} = \left\{ (x, y) : 0 \leq x \leq \sqrt{f\left(\frac{y}{x}\right)} \right\} \subset \mathbb{R}^2 \quad (1.4.1)$$

où $f \geq 0$ est une fonction intégrable.

Si (X, Y) est un vecteur aléatoire uniformément distribué sur \mathcal{D} , alors

$$Z = \frac{Y}{X} \quad \text{a une densité } f \text{ souhaitée.}$$

Démonstration

Considérons la transformation

$$T : (X, Y) \longrightarrow (W, Z) \text{ telle que}$$

$$z = \frac{y}{x}, \quad w = x \quad \text{donc} \quad y = zw. \quad (1.4.2)$$

Le jacobien de transformation est égal à z . La densité jointe de (W, Z) est

$$g(w, z) = \begin{cases} w/|\mathcal{D}|, & \text{si } 0 \leq z \leq f^{1/2}(z), \\ 0, & \text{sinon} \end{cases}. \quad (1.4.3)$$

La densité marginale de Z est

$$\int_0^{f^{1/2}(z)} w/|\mathcal{D}| dw = \left[f^{1/2}(z) \right]^2 / 2|\mathcal{D}| = f(z)/2|\mathcal{D}|. \quad (1.4.4)$$

On note que le chiffre 2 doit apparaître dans l'équation (1.4.3) de sorte que la densité marginale de l'équation (1.4.3) puisse être intégrable à 1 ce qui implique que l'aire de \mathcal{D} est toujours égale à 1/2.

Remarque

Pour toute fonction non négative h avec intégrale K finie, soit

$$\mathcal{D}_h = \left\{ (x, y) : 0 \leq x \leq \sqrt{h\left(\frac{y}{x}\right)} \right\}.$$

Par un argument similaire à celui ci-dessus, si (X, Y) est uniformément distribué sur \mathcal{D}_h , alors $Z = Y/X$ a une densité $f(z) = K^{-1}h(z)$ et la mesure de $\mathcal{D}_h = K/2$.

Dans beaucoup de cas l'équation $x = h^{1/2}(z)$ peut être résolue (typiquement pour y en fonction de x) et la frontière de \mathcal{D}_h peut être trouvée explicitement. Les exemples suivants illustrent ce point, aussi bien que la méthode de base.

Exemple 1.4.1. (*Uniforme*)

$$\text{Soit } f(z) = \begin{cases} (b-a)^{-1}, & \text{si } z \in (a, b) \\ 0, & \text{sinon} \end{cases}$$

$$\text{alors } \mathcal{D} = \{0 \leq x \leq (b-a)^{-1/2}, a \leq y/x \leq b\}$$

ce qui est un triangle avec des sommets $(0, 0)$, $(\lambda, a\lambda)$, $(b\lambda, \lambda)$ pour $\lambda = (b-a)^{-1/2}$.

Exemple 1.4.2. (*Normale*)

$$\text{Soit } r(z) = \exp(-z^2/2), \text{ alors}$$

$$\mathcal{D}_h = \{0 \leq x \leq \exp(-y^2/(4x^2))\} = \{y^2 \leq -4x^2 \ln x\}.$$

La frontière est la courbe $y = \pm 2x(-\ln x)^{1/2}$, $0 \leq x \leq 1$.

Dans la suite, nous attacherons plus d'importance à toutes les conditions qui nous assurent que $\mathcal{D} \subseteq [0, b) \times [a_-, a_+]$ pour des constantes finies $b \geq 0$, $a_- \leq 0$ et $a_+ \geq 0$, avec $b = \sup \sqrt{f(z)}$, $a_- = \inf z \sqrt{f(z)}$ et $a_+ = \sup z \sqrt{f(z)}$.

Il est à noter que la frontière de \mathcal{D} peut être trouvée par $\{(x(z), y(z)) : z \in \mathbb{R}\}$ où

$$x(z) = \sqrt{f(z)} \tag{1.4.5}$$

$$y(z) = z \sqrt{f(z)}. \tag{1.4.6}$$

Alors, \mathcal{D} sera enfermée dans le rectangle $\mathcal{B} \subseteq [0, b) \times [a_-, a_+]$ si et seulement si

- (1) $f(z)$ est bornée,
- (2) $z^2 f(z)$ est bornée.

Fondamentalement, ceci inclut les densités bornées telles que la normale, la gamma, la bêta et l'exponentielle.

Pour plus de simplicité nous allons traiter les densités qui satisfassent les conditions (1) et (2)

Algorithme

- (1) Trouver b, a_- et a_+ pour le rectangle \mathcal{B} . On note que $b \geq \sup \sqrt{f(z)}$,
 $a_- \leq \inf z \sqrt{f(z)}$ et $a_+ \geq \sup z \sqrt{f(z)}$.
- (2) Générer X uniformément sur $[0, b)$, et Y uniformément sur $[a_-, a_+]$.
- (3) Poser $Z = Y/X$.
- (4) Si $X^2 \leq f(Z)$ alors on garde Z , sinon retourner à l'étape (2).

Le nombre prévu d'itérations est de

$$\frac{b(a_+ - a_-)}{m(\mathcal{D})} = \frac{2b(a_+ - a_-)}{\int_{-\infty}^{+\infty} f(z) dz} \quad (1.4.7)$$

qu'on appellera constante de rejet.

Les densités les plus intéressantes sont celles pour lesquelles la majeure partie de \mathcal{D} est enfermée dans le rectangle \mathcal{B} .

Pour mieux s'assurer de la rapidité de notre algorithme, on peut introduire une ou plusieurs étapes d'acceptation et rejet.

L'algorithme sera donc de la forme :

- (1) Trouver b, a_- et a_+ pour le rectangle \mathcal{B} , on note que $b \geq \sup \sqrt{f}(z)$,
 $a_- \leq \inf z\sqrt{f}(z)$ et $a_+ \geq \sup z\sqrt{f}(z)$.
- (2) Générer X uniformément sur $[0, b]$, et Y uniformément sur $[a_-, a_+]$.
- (3) Poser $Z = Y/X$.
- (4) Si $X^2 \leq f(Z)$ alors on garde Z , sinon retourner à l'étape (2).

Exemple 1.4.3.

Cet exemple est basé sur la distribution d'une variable aléatoire de loi χ , la loi de la valeur absolue d'une normale. On considère $a_1 = 0$ et $a_2 = 1$.

Ainsi donc

$$\mathcal{D} = \left\{ (x, y) : 0 \leq x \leq \exp((y/2x)^2) \right\},$$

$$r(z) = \exp(-(z^2/2)), z > 0,$$

$$m(\mathcal{D}) = \frac{1}{2} \int_0^\infty \exp(-z^2/2) dz = \frac{1}{2} \sqrt{\pi/2},$$

$$x^* = 1, \quad y_* = 0 \quad y^* = \sqrt{2/e}.$$

Par la formule

$$m(\mathcal{D})/m(\mathcal{B}) = \sqrt{\pi e}/4 = 0,7306.$$

Le test d'acceptation dans l'algorithme est de la forme

$$X^2 \leq e^{(Y/X)^2/2},$$

lequel est équivalent à l'événement suivant :

$$Y^2 \leq -4X^2 \ln X.$$

La probabilité d'acceptation est

$$P = m(\mathcal{D})/m(\mathcal{B}) = \sqrt{(\pi e)}/4 = 0,7306.$$

Exemple 1.4.4. (*Normale toute entière*)

$r(z)$	$e^{\frac{-z^2}{2}}, (z \in \mathbb{R})$
$b = \sup \sqrt{r(z)}$	1
$a_+ = \sup z\sqrt{r(z)}, a_- = \inf z\sqrt{r(z)}$	$\sqrt{(\frac{2}{e})}, -\sqrt{(\frac{2}{e})}$
$m(\mathcal{D}), m(\mathcal{B})$	$\sqrt{\pi/2}, 2\sqrt{e/2}$
Constante de rejet	$\frac{4}{\pi e}$
Condition d'acceptation	$z^2 \leq -4 \log x$
Condition d'acceptation rapide	$z^2 \leq 4(-cx + 1 + \log c) \quad (c > 0)$ $z^2 \leq 4 - 4x$ $z^2 \leq 6 - 8x + 2x^2$
Condition de rejet rapide	$z^2 > 4(\frac{c}{x} - 1 - \log c) \quad (c > 0)$ $z^2 \geq \frac{4}{x} - 4$ $z^2 \geq \frac{2}{x} - 2x$

L'idée d'avoir une condition d'acceptation rapide et une condition de rejet rapide est de rendre les calculs plus rapides. Typiquement, accepter avec la condition d'acceptation rapide signifie qu'on aurait accepté avec la condition d'acceptation de base et rejeter avec la condition de rejet rapide signifie qu'on rejetterait avec la condition d'acceptation de base. Ainsi, l'utilisateur teste si on accepte avec la condition d'acceptation rapide ou si on rejette avec la condition de rejet rapide. Si l'une ou l'autre de ces conditions est vérifiée alors on rend la décision. Sinon, on doit revenir avec la vérification de la condition d'acceptation de base qui prend plus de temps à faire.

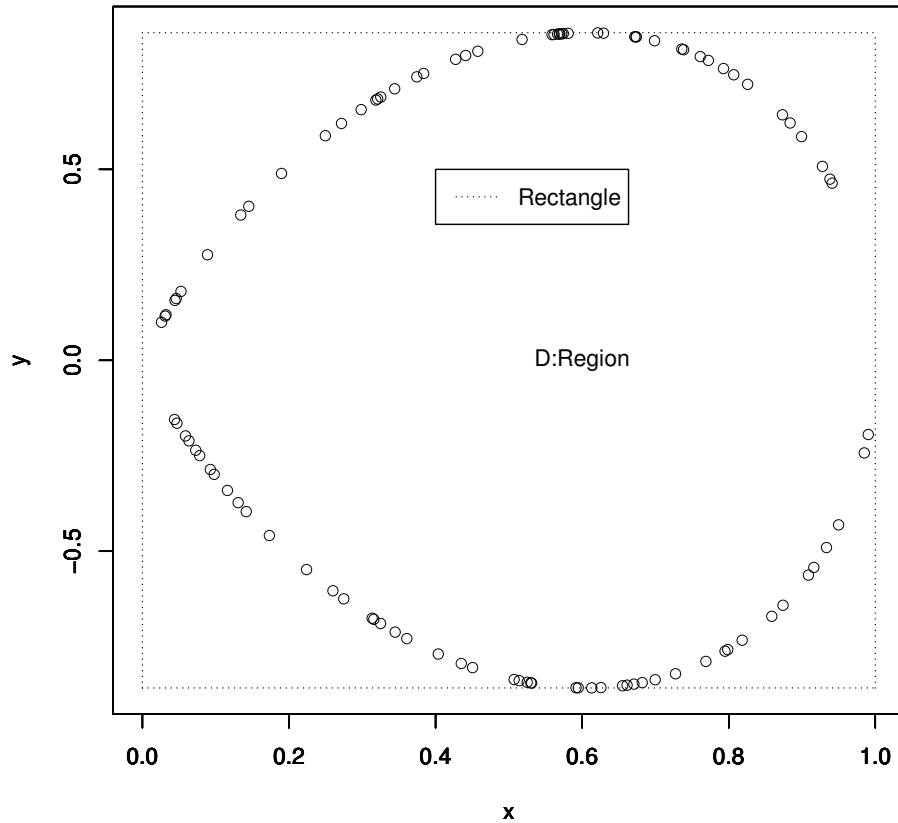


FIGURE 1.2. Région d'acceptation \mathcal{D} d'une normale obtenue par la méthode du quotient d'uniformes

Quelques généralisations

1) Si $f(z) = f_0(z)f_1(z)$ et le vecteur (Y, X) a une densité proportionnelle à $f_0(y/x)$ lorsqu'on évalue en un point (y, x) sur

$$\mathcal{D} = \left\{ (y, x) : 0 < x < \sqrt{(f_1(y/x))} \right\}.$$

Alors $Z = Y/X$ a une densité proportionnelle à f .

1.4.2. Cas multivarié

1.4.2.1. Méthode et principe général

Supposons un vecteur aléatoire $X = (X_1, \dots, X_n)'$ dont la fonction densité est de la forme

$$f(x_1, x_2, \dots, x_n) = g(x_1, x_2, \dots, x_n)/K, \quad (1.4.8)$$

où $g(x) = g(x_1, x_2, \dots, x_n) \geq 0$, pour tout $x \in \mathbb{R}^n$ et

$$K = \int_{\mathbb{R}^n} g(t) dt. \quad (1.4.9)$$

Soit $H \subset \mathbb{R}^{n+1}$ un domaine borné tel que

$$H = \{(v_0, v_1, \dots, v_n)' : h(v_0, v_1, \dots, v_n) \leq 0, v_0 > 0\} \quad (1.4.10)$$

où la fonction $h : \mathbb{R}_+ \times \mathbb{R}^n \longrightarrow \mathbb{R}$ est telle que

$$h(v_0, v_1, \dots, v_n) = \ln(v_0) - \left(p/(n+p) \right) \ln \left(g(\varphi(v_0, v_1, \dots, v_n)) \right) \leq 0$$

avec $\varphi : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}^n$ la transformation donnée par

$$\varphi(v) = \varphi(v_0, v_1, \dots, v_n) \quad (1.4.11)$$

$$= \left(a_1 + \frac{v_1}{\sqrt[p]{v_0}} + \dots + a_n + \frac{v_n}{\sqrt[p]{v_0}} \right)', p > 0. \quad (1.4.12)$$

Notons que la transformation φ est la généralisation de la transformation

$$\varphi_0(v_0, v_1, \dots, v_n) = \left(\frac{v_1}{\sqrt{v_0}}, \dots, \frac{v_n}{\sqrt{v_0}} \right)'$$

Théorème 1.4.3. *Si le vecteur aléatoire $V = (V_0, V_1, \dots, V_n)'$ est uniformément distribué dans le domaine H , alors la fonction de densité du vecteur aléatoire $X = \varphi(V) = \varphi(V_0, V_1, \dots, V_n)$ est de la forme (1.4.8)*

Démonstration

Par (1.4.11) nous avons

$$H = \left\{ v = (v_0, \dots, v_n)' \mid v \in \mathbb{R}_+ \times \mathbb{R}^n, 0 < v_0 \leq g^{p/(n+p)}(\varphi(v)) \right\}. \quad (1.4.13)$$

Considérons le changement de variables suivant :

$$\begin{cases} x_0 = v_0, \\ x_i = a_i + v_i / \sqrt[p]{v_0}, \quad 1 \leq i \leq n \end{cases} \quad (1.4.14)$$

Notons que cette transformation est biunivoque de H dans $H^* \subset \mathbb{R}_+ \times \mathbb{R}^n$ où

$$H^* = \left\{ (x_0, x_1, \dots, x_n)' : 0 < x_0 \leq g^{p/(n+p)}(x_1, \dots, x_n) \right\} \quad (1.4.15)$$

et l'inverse devient

$$\begin{cases} v_0 = x_0, \\ v_i = (x_i - a_i) \sqrt[p]{x_0}, \quad 1 \leq i \leq n. \end{cases} \quad (1.4.16)$$

Donc le jacobien est

$$\frac{D(v_0, v_1, \dots, v_n)}{D(x_0, x_1, \dots, x_n)} = \left(\sqrt[p]{x_0} \right)^n. \quad (1.4.17)$$

En utilisant ce fait, par un simple calcul nous obtenons

$$m(H) = \int \cdots \int_H dv_0 dv_1 \cdots dv_n = pK/(n+p), \quad (1.4.18)$$

car

$$m(H) = \int \cdots \int_H dv_0 dv_1 \cdots dv_n = \left[\int \cdots \int_{\mathbb{R}} \left(\int_0^{g^{p/(p+n)}(x_1, \dots, x_n)} (\sqrt[p]{x_0})^n dx_0 \right) dx_1 \cdots dx_n \right]$$

et

$$\int (\sqrt[p]{x_0})^n dx_0 = \frac{x_0^{\frac{n}{p}+1}}{\frac{n}{p}+1}.$$

Donc

$$\int_0^{g^{p/(p+n)}(x_1, \dots, x_n)} (\sqrt[p]{x_0})^n dx_0 = \frac{p}{n+p} g(x_1, \dots, x_n).$$

Or

$$f(x_1, x_2, \dots, x_n) = g(x_1, x_2, \dots, x_n)/K$$

et f est une densité, donc

$$m(H) = \frac{p}{n+p} \left[\int \cdots \int_{\mathbb{R}} K f(x_1, x_2, \dots, x_n) dx_1 \cdots dx_n \right] = \frac{p}{n+p} K.$$

Puisque notre vecteur aléatoire V est uniformément distribué dans H , sa fonction densité est donnée par :

$$u(v_0, \dots, v_n) = \begin{cases} 1/m(H) & \text{si } (v_0, v_1, \dots, v_n)' \in H \\ 0, & \text{sinon.} \end{cases} \quad (1.4.19)$$

Notons la fonction de répartition cumulative de X par

$$F(b_1, b_2, \dots, b_n) = P(X_1 \leq b_1, X_2 \leq b_2, \dots, X_n \leq b_n). \quad (1.4.20)$$

Ainsi,

$$F(b_1, b_2, \dots, b_n) = \left[\int \cdots \int_{H \cap W} dv_0 dv_1 \cdots dv_n \right] / m(H) \quad (1.4.21)$$

où

$$W = \left\{ (v_0, v_1, \dots, v_n)' : a_i + v_i / \sqrt[p]{v_0} < b_i, 1 \leq i \leq n, 0 < v_0 < g^{p/(p+n)}((\varphi(v_0, v_1, \dots, v_n))) \right\}.$$

Par les expressions (1.4.14), (1.4.16) et (1.4.12) nous obtenons

$$F(b_1, b_2, \dots, b_n) = \left[\int_{-\infty}^{b_1} \cdots \int_{-\infty}^{b_n} \left(\int_{-\infty}^{g^{p/(p+n)}(x_1, \dots, x_n)} (\sqrt[p]{x_0})^n dx_0 \right) dx_1 \cdots dx_n \right] / m(H).$$

Finalement, on trouve

$$F(b_1, b_2, \dots, b_n) = \left[\int_{-\infty}^{b_1} \cdots \int_{-\infty}^{b_n} g(x_1, \dots, x_n) dx_1 \cdots dx_n \right] / K. \quad (1.4.22)$$

Puisque (1.4.22) est la fonction de répartition correspondant à celle de f donnée par (1.4.8), le théorème est démontré.

Ce théorème induit l'algorithme général suivant pour générer le vecteur aléatoire X .

1.4.2.2. Algorithmes

Algorithme (A)

- (1) Faire les initialisations appropriées.
- (2) Générer le vecteur aléatoire V uniformément sur le domaine borné H .
- (3) Poser $X = \varphi(V)$.

Pour mettre en application cet algorithme dans le cas de la fonction de densité particulière donnée par (1.4.8), deux problèmes importants doivent d'abord être résolus, à savoir :

- (1) Vérifier si le domaine H est borné.
- (2) Construire un algorithme pour générer V sur H .

Pour mettre en application ces résultats nous allons chercher un pavé I de la forme

$$I = [v_0^{(1)}, v_0^{(2)}] \times [v_1^{(1)}, v_1^{(2)}] \times \cdots \times [v_n^{(1)}, v_n^{(2)}] \quad (1.4.23)$$

$$0 \leq v_0^{(1)} < v_0^{(2)} < \infty, -\infty < v_i^{(1)} < v_i^{(2)} < \infty, 1 \leq i \leq n$$

tel que $H \subset I \subset \mathbb{R}_+ \times \mathbb{R}^n$. Si on parvient à trouver ce pavé I alors l'algorithme (A) est spécifié par l'algorithme de rejet suivant :

Algorithme GENUNV

- Etape 1.1 Générer des variables aléatoires uniformes et indépendantes U_0, U_1, \dots, U_N sur l'intervalle $[0, 1]$.

Calculer $v_i^* = v_i^{(1)} + (v_i^{(2)} - v_i^{(1)})U_i$, $0 \leq i \leq n$ et poser $V^* = (V_0^*, V_1^*, \dots, V_n^*)$.

- Etape 1.2 Si $V^* \notin H$ retourner à l'étape (1.1). (rejet de V^*)

- Etape 1.3 Si V^* est accepté alors garder $V = V^*$.

On note que la performance de l'algorithme A dépend de la performance de l'algorithme GENUNV lequel est caractérisé par la probabilité d'acceptation dans l'étape 1.2. Cette probabilité est

$$p_a = m(H) / \prod_{i=0}^n [v_i^{(2)} - v_i^{(1)}] \quad (1.4.24)$$

et en utilisant (1.4.18) nous avons

$$p_a = p_a(n, p) = pK / \left[(n + p) \prod_{i=0}^n (v_i^{(2)} - v_i^{(1)}) \right]. \quad (1.4.25)$$

Un bon algorithme doit avoir p_a très élevé, et ceci peut être obtenu seulement si le pavé I donné par (1.4.23) est le plus petit pavé contenant H .

Pour déterminer ce pavé I nous devons résoudre le problème d'optimisation suivant

$$v_i^{(1)} = \min_{v \in H} v_i, \quad v_i^{(2)} = \max_{v \in H} v_i, \quad v = (v_0, v_1, \dots, v_n)', \quad 0 \leq i \leq n. \quad (1.4.26)$$

En appliquant la méthode des multiplicateurs de Lagrange au problème (1.4.26), nous aurons

$$\min_{(v, \lambda) \in H \times \mathbb{R}} F_i(v, \lambda), \quad \max_{(v, \lambda) \in H \times \mathbb{R}} F_i(v, \lambda), \quad 0 \leq i \leq n.$$

$$F_i(v, \lambda) = v_i + \lambda h(v), \quad \lambda \in \mathbb{R}. \quad (1.4.27)$$

Pour un i fixé, $0 \leq i \leq n$, notons par $(^*v^{(1)}, \lambda^{(1)})$, $(^*v^{(2)}, \lambda^{(2)})$ les points optimaux dans \mathbb{R}^{n+1} de ces deux problèmes d'optimisation de (1.4.27) respectivement. Les bornes du pavé I sont

$$v_i^{(1)} = ^*v^{(1)}, \quad v_i^{(2)} = ^*v^{(2)}. \quad (1.4.28)$$

Remarque :

- (1) Les constantes $a_i, i = 1, 2, \dots, n$ définies à partir de la transformation $\varphi : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}^n$ avec $\varphi(V) = \left(a_1 + \frac{v_1}{\varphi/v_0}, \dots, a_n + \frac{v_n}{\varphi/v_0} \right)', p > 0$, nous permettent de définir le plus petit pavé contenant H . Donc les bonnes densités sont les densités dans lesquelles H occupe presque toute la place dans le pavé.

- (2) Dans le théorème (1.4.3), la transformation φ donnée par (1.4.27) implique l'existence du vecteur constant (a_1, a_2, \dots, a_n) . Si on considère les valeurs particulières de ce vecteur et de p on obtient les diverses formes pour φ qui ont été considérées.

Cas général

Si V, U sont uniformes sur

$$\mathcal{C} = \left\{ (v, u) : v \in \mathbb{R}^d, 0 < U < \sqrt[d+1]{f(v/u + \mu)} \right\}$$

alors $X = V/U + \mu$ a une densité proportionnelle à f .

Définition 1.4.1. *L'algorithme $A(n, p_0)$ est le plus efficace dans la famille*

$$\{A(n, p), n \in (N)_+, p > 0\} \quad \text{si} \quad p_a(n, p_0) = \max\{p_a(n, p), p > 0\}.$$

L'algorithme $A(n, p)$ le plus efficace pour des distributions particulières

Distribution Normale Multivariée

Le vecteur aléatoire $X = (X_1, \dots, X_n)$ de dimension n de distribution $N(\mu, \Sigma)$ et sa fonction densité f est

$$f(x) = \exp(-Q(X)/2) / K$$

avec

$$Q(x) = (x - \mu)' \Sigma^{-1} (x - \mu), \quad K = (2\pi)^{n/2} (\det \Sigma)^{1/2} \quad (1.4.29)$$

où μ est le vecteur moyen de X et Σ est la matrice de covariance symétrique et définie positive.

Pour générer un vecteur aléatoire selon $N(\mu, \Sigma)$ il suffit de construire un algorithme de génération de ce vecteur selon $N(0, \Sigma)$.

Dans la suite nous allons supposer $\mu = 0$ dans (1.4.29) et on notera

$$A = \Sigma^{-1} = (a_{ij})_{1 \leq i \leq n}, \Sigma = (\sigma_{ij})_{1 \leq i \leq n}. \quad (1.4.30)$$

Proposition 1.4.1. *L'algorithme le plus efficace dans la famille $\{A(n, p), p > 0\}$ pour générer un vecteur aléatoire de X selon $N(0, \Sigma)$ est l'algorithme $A(n, 2)$.*

Démonstration

De (1.4.8) il découle que la fonction $h(v_0, v_1, \dots, v_n)$ impliquée dans la définition du domaine H est

$$h(v_0, v_1, \dots, v_n) = \ln(v_0) + \left[p/(2(n+p)) \right] \sum_{i=1}^n \sum_{j=1}^n a_{ij} v_i v_j / v_0^{2/p}. \quad (1.4.31)$$

Afin de trouver un pavé I sous la forme requise par l'algorithme GENUNV nous devons résoudre les problèmes d'optimisation avec la méthode de Lagrange.

$$F_s(v_0, v_1, \dots, v_n; \lambda) = v_s + \lambda h(v_0, v_1, \dots, v_n) \quad (1.4.32)$$

$$\lambda \neq 0, 0 \leq s \leq n, v_0 > 0.$$

Donc les limites $v_s^{(1)}, v_s^{(2)}$ lesquelles définissent le pavé I doivent satisfaire le système d'équations suivant :

$$\begin{cases} \partial F_s(\bar{v}_0, \bar{v}_1, \dots, v_s^j, \dots, \bar{v}_n; \lambda) / \partial v_r = 0 \\ \partial F_s(\bar{v}_0, \bar{v}_1, \dots, v_s^j, \dots, \bar{v}_n; \lambda) / \partial \lambda = 0 \end{cases} \quad (1.4.33)$$

$$0 \leq s \leq n, 0 \leq r \leq n, j = 1, 2.$$

Pour (1.4.33), le système pour $s = 0$ est

$$\begin{aligned} 1 + (\lambda/v_0) \left[1 - \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij} v_i v_j \right) / (v_0^{2/p} (n+p)) \right] &= 0 \\ \lambda(p/(n+p)) \left[\left(\sum_{i=1}^n a_{ir} v_i \right) / v_0^{2/p} \right] &= 0, \quad 0 \leq r \leq n \\ \ln(v_0) + p \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij} v_i v_j \right) / \left[2(n+p) v_0^{2/p} \right] &= 0. \end{aligned} \quad (1.4.34)$$

Multiplions la seconde équation de (1.4.34) par v_r et sommons pour $r = 1, \dots, n$ ce qui nous donne :

$$\sum_{i=1}^n \sum_{r=1}^n a_{ir} v_i v_r = 0. \quad (1.4.35)$$

En utilisant la relation dans la dernière équation (1.4.34) nous obtenons :

$$\ln(v_0) = 0. \quad (1.4.36)$$

Ce qui donne finalement

$$v_0^2 = 1. \quad (1.4.37)$$

Avec $v_0 > 0$, nous pouvons prendre

$$v_0^1 = 0. \quad (1.4.38)$$

Afin de trouver v_s^1, v_s^2 , pour s fixé, $0 \leq s \leq n$ nous devons résoudre le système suivant de $n+2$ équations dérivées de (1.4.33)

$$\begin{aligned} (\lambda/v_0) \left[1 - \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij} v_i v_j \right) / (v_0^{2/p} (n+p)) \right] &= 0, \\ \delta_{sr} \lambda(p/(n+p)) \left[\left(\sum_{i=1}^n a_{ir} v_i \right) / v_0^{2/p} \right] &= 0, \quad 0 \leq r \leq n \\ \ln(v_0) + p \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij} v_i v_j \right) / \left[2(n+p) v_0^{2/p} \right] &= 0, \end{aligned} \quad (1.4.39)$$

où

$$\delta_{sr} = \begin{cases} 1 & \text{if } r = s \\ 0 & \text{if } r \neq s. \end{cases} \quad (1.4.40)$$

Si on note

$$w = p \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij} v_i v_j \right) / \left[2(n+p) v_0^{2/p} \right], \quad (1.4.41)$$

alors de la première et de la dernière formule (1.4.39) nous obtenons

$$(\lambda/v_0) [1 - 2w/p] = 0, \ln(v_0) + w = 0. \quad (1.4.42)$$

Avec $v_0 > 0$ et $\lambda \neq 0$ nous avons

$$w = p/2, v_0 = \exp(-p/2). \quad (1.4.43)$$

Multiplions la seconde formule de (1.4.39) par v_r , $0 \leq r \leq n$. On obtient par addition

$$\lambda = v_s/p. \quad (1.4.44)$$

Comme $A = \Sigma^{-1}$ nous avons

$$\sum_{i=1}^n a_{ij} \sigma_{js} = \delta_{is}, 0 \leq s \leq n. \quad (1.4.45)$$

Si on multiplie encore la seconde formule (1.4.39) par σ_{rs} , $0 \leq r \leq n$, et en effectuant l'addition sur r et en tenant compte de (1.4.40), (1.4.41) et (1.4.42),

nous obtenons

$$v_s^{(1)} = (n + p)\sigma_{ss}/e, 0 \leq s \leq n, \quad (1.4.46)$$

et donc

$$v_s^{(1)} = -\left((n + p)\sigma_{ss}/e\right)^{1/2}, v_s^{(2)} = \left((n + p)\sigma_{ss}/e\right)^{1/2}, \quad s = 1, \dots, n. \quad (1.4.47)$$

La probabilité d'acceptation donné par (1.4.24), (en utilisant (1.4.37), (1.4.38) et (1.4.47)), est

$$p_a(n, p) = \left[p/(n + p)^{n/2+1} \right] \left((\pi e/2)^{n/2} (\det \sigma) / (\sigma_{11} \dots \sigma_{nn}) \right)^{1/2}. \quad (1.4.48)$$

Le maximum de cette probabilité est obtenu pour $p = 2$ et donc la proposition est démontrée.

1.5. MÉTHODE DE MONTE CARLO

1.5.1. Introduction

Hammersley (1956) définit les méthodes de Monte Carlo comme étant la mise en oeuvre de la branche des mathématiques qui s'intéresse aux expériences sur des nombres aléatoires.

Chaque fois qu'on déduit des lois générales d'observations particulières, les conclusions sont entachées d'incertitude dans la mesure où ces observations particulières sont seulement un échantillon plus ou moins représentatif de la totalité des observations qui auraient pu être faites. Dans ce cas, les réponses de Monte Carlo deviennent incertaines, mais elles restent fructueuses si on arrive à rendre l'incertitude pratiquement négligeable.

La façon la plus naturelle de réduire cette incertitude est de réunir davantage d'observations pour la fonder. Pour éviter ce travail d'expérimentation excessif

ou même impossible, on peut avec profit changer ou du moins modifier le problème initial de telle façon que l'incertitude des réponses soit diminuée.

De tels procédés sont connus sous le nom de technique de réduction de la variance. Ici, l'incertitude se mesure en sommant le biais au carré et la variance.

1.5.2. Efficacité

Dans les travaux de Monte Carlo l'objectif principal est d'obtenir une erreur-type raisonnablement faible dans le résultat final.

L'efficacité d'un processus de Monte Carlo peut être considérée comme inversement proportionnelle au produit de la variance d'échantillonnage et de la quantité du travail effectuée pour obtenir l'estimation.

Ainsi, augmenter le travail peut rapporter de bons résultats si cela amène à une grande diminution de la variance échantillonnale.

1.5.3. Intégration de Monte Carlo

Dans cette section nous allons donner deux techniques pour évaluer l'intégrale unidimensionnelle suivante :

$$\int_a^b f(x)dx$$

par la méthode de Monte Carlo.

1.5.3.1. Méthode de Monte Carlo de « Noir ou Blanc »

Considérons le problème de calculer l'intégrale ci-haut citée et pour plus de simplicité, nous allons supposer que la fonction f est bornée :

$$0 \leq f(x) \leq c \quad \text{pour tout } x \in [a, b]. \quad (1.5.1)$$

Soit Ω le rectangle suivant :

$$\Omega = \left\{ (x, y) : a \leq x \leq b \quad , \quad 0 \leq y \leq c \right\}. \quad (1.5.2)$$

Soit (X, Y) un vecteur aléatoire uniformément distribué sur le rectangle précédant avec la fonction densité

$$f_{XY}(x, y) = \begin{cases} \frac{1}{c(b-a)} & \text{si } (x, y) \in \Omega \\ 0 & \text{sinon} \end{cases} \quad (1.5.3)$$

Notons $S = \{(x, y) : y \leq f(x)\}$.

Soit $p = P[(X, Y) \in S]$ la probabilité que le vecteur aléatoire soit sous la courbe f , (voir la figure 1.3 pour une illustration.)

La surface sous la courbe f est donnée par $m(S)$ et

$$m(S) = \int_a^b f(x) dx.$$

On obtient

$$p = \frac{m(S)}{m(\Omega)} \quad (1.5.4)$$

$$= \frac{\int_a^b f(x) dx}{c(b-a)} \quad (1.5.5)$$

$$= \frac{I}{c(b-a)}. \quad (1.5.6)$$

Supposons qu'on génère n vecteurs aléatoires indépendants $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$. Le paramètre p sera estimé par

$$\hat{p} = \frac{n_h}{n}, \quad (1.5.7)$$

où n_h est le nombre de fois que l'événement $f(X_i) \geq Y_i$, $i = 1, 2, \dots, n$ se réalise, c'est-à-dire le nombre de «Noir» et $n - n_h$ est le nombre de «Blanc».

Par les expressions (1.5.6) et (1.5.7) l'intégrale I peut être estimée par :

$$I \approx \hat{\theta}_{HM} = c(b-a) \frac{n_h}{n}.$$

Puisque chacune des n épreuves constitue une épreuve de Bernoulli de probabilité p de «Noir»

$$\begin{aligned} E(\hat{\theta}_{HM}) &= c(b-a)E\left(\frac{n_h}{n}\right) = c(b-a)\frac{E(n_h)}{n} \\ &= pc(b-a) = I. \end{aligned}$$

Ainsi, $\hat{\theta}_{HM}$ est un estimateur sans biais de I .

La variance de \hat{p} sera égale à

$$\begin{aligned} \text{Var}(\hat{p}) &= \text{Var}\left(\frac{n_h}{n}\right) = \frac{1}{n^2}\text{Var}(n_h) = \frac{1}{n}p(1-p), \\ &= \frac{1}{n} \frac{I}{[c(b-a)]^2} [c(b-a) - I]. \end{aligned}$$

$$\text{Var}(\hat{\theta}_{HM}) = [c(b-a)]^2 \text{Var}(\hat{p}), \quad (1.5.8)$$

$$= [c(b-a)]^2 \frac{1}{n} p(1-p), \quad (1.5.9)$$

$$= \frac{I}{n} [c(b-a) - I]. \quad (1.5.10)$$

On peut se demander quelle taille d'échantillon n avons-nous besoin pour que le résultat de la méthode de «Noir» ou «Blanc» puissent être assez précis.

Théorème 1.5.1. *Pour que cette méthode de «Noir» ou «Blanc» puisse être assez précise on se donne une valeur de ϵ qui fixe le degré de précision et on constate qu'il suffit d'avoir*

$$n \geq \frac{(1-p)p[c(b-a)]^2}{(1-\alpha)\epsilon^2} \quad \text{où } \alpha \in (0, 1) \quad \text{et } \epsilon > 0.$$

Démonstration

Une question essentielle qu'on doit se poser consiste à quantifier la précision de notre estimateur. Ainsi, donnons nous $\alpha \in (0, 1)$ et $\epsilon > 0$ et cherchons à savoir

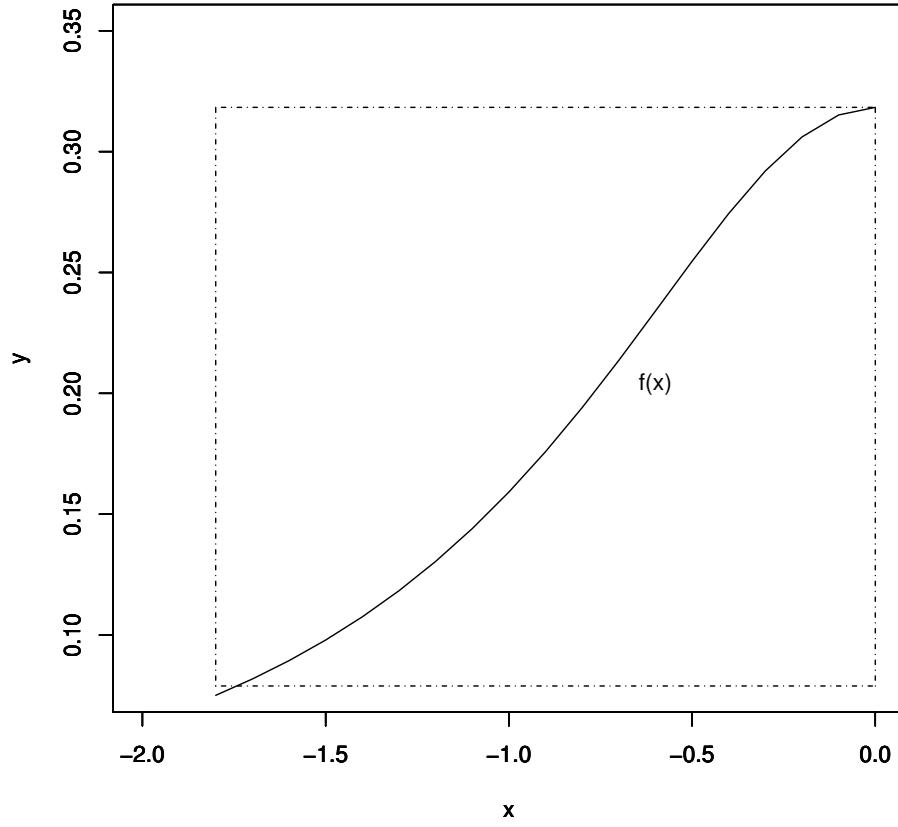


FIGURE 1.3. Graphique : Méthode de Noir ou Blanc où $a = -1,7$ et $b = 0$

quelle taille d'échantillon dont on a besoin pour nous garantir le résultat suivant :

$$P[|\hat{\theta}_{HM} - I| \leq \varepsilon] \geq \alpha. \quad (1.5.11)$$

Par l'inégalité de Chebyshev on a :

$$P[|\hat{\theta}_{HM} - I| \leq \varepsilon] \geq 1 - \frac{\text{Var}(\hat{\theta}_{HM})}{\varepsilon^2}. \quad (1.5.12)$$

De (1.5.12) avec (1.5.11) nous avons

$$\alpha \leq 1 - \frac{\text{Var}(\hat{\theta}_{HM})}{\varepsilon^2}. \quad (1.5.13)$$

Par la substitution de (1.5.10) dans (1.5.13), nous obtenons

$$\alpha \geq \frac{p(1-p)[c(b-a)]^2}{n\varepsilon^2}.$$

En isolant la valeur de n nous avons

$$n \geq \frac{p(1-p)[c(b-a)]^2}{(1-\alpha)\varepsilon^2}.$$

Ce qui devient une borne supérieure sur la taille d'échantillon minimale pour que (1.5.11) puisse satisfaire le degré de précision qu'on se donne.

1.5.3.2. *Algorithme*

Algorithme de la méthode de Noir ou Blanc de Monte Carlo

- (1) Générer une suite $\left\{U_i\right\}_{i=1}^{2n}$ nombres aléatoires de loi $\mathcal{U}(0, 1)$.
- (2) Calculer $X_i = a + U_i(b-a)$ et $f(X_i)$ $i = 1, 2, \dots, n$.
- (3) Compter le nombre de cas n_h pour lequel $f(X_i) \geq cU_{i+n}$.
- (4) Estimer l'intégrale I par

$$\hat{\theta}_{HM} = c(b-a) \frac{n_h}{n}.$$

1.5.3.3. *Méthode de Monte Carlo moyenne échantillonnale*

Une autre façon de calculer l'intégrale

$$I = \int_a^b f(x)dx \tag{1.5.14}$$

est d'écrire I sous forme d'une espérance de variable aléatoire.

Donc on peut réécrire l'intégrale I de la façon suivante :

$$I = \int_a^b \frac{f(x)}{g(x)} g(x) dx.$$

En supposant que g est une densité telle que $g(x) > 0$ quand $f(x) \neq 0$ alors on a

$$I = E \left[\frac{f(X)}{g(X)} \right],$$

où X est une variables aléatoire distribuée selon une loi de densité g .

Considérons un échantillon X_1, \dots, X_n distribué selon une loi de densité g donc l'estimateur de I sera

$$\hat{\theta}_{Me} = \frac{1}{n} \sum_i^n \frac{f(X_i)}{g(X_i)}.$$

Cet estimateur est sans biais. La variance de cet estimateur est égale à

$$\begin{aligned} \text{Var}(\hat{\theta}_{Me}) &= \frac{1}{n} \text{Var} \left[\frac{f(X)}{g(X)} \right], \\ &= \frac{1}{n} \left[\int \frac{f^2(x)}{g(x)} dx - I^2 \right]. \end{aligned} \quad (1.5.15)$$

1.5.3.4. *Algorithme*

Algorithme de Monte Carlo moyenne échantillonnale

- (1) Générer une suite de n variables aléatoires $\left\{ X_i \right\}_{i=1}^n$ selon une loi de densité g .
- (2) Calculer $g(X_i)$ et $f(X_i)$ pour $i = 1, \dots, n$.
- (3) Calculer la moyenne échantillonnale $\hat{\theta}_{Me}$ et $\text{Var}(\hat{\theta}_{Me})$.

Pour plus de simplicité supposons que

$$g_X(x) = \begin{cases} \frac{1}{(b-a)} & \text{si } a < x < b, \\ 0 & \text{si sinon.} \end{cases}$$

Ainsi,

$$E[f(X)] = \frac{I}{b-a},$$

et

$$I = (b-a)E[f(X)].$$

L'estimateur sans biais de I est sa moyenne échantillonnale

$$\hat{\theta}_{Me} = (b-a) \frac{1}{n} \sum_{i=1}^n f(X_i).$$

La variance de $\hat{\theta}_{Me}$

$$\text{Var}(\hat{\theta}_{Me}) = \text{Var}\left[\frac{1}{n}(b-a) \sum_{i=1}^n f(X_i)\right], \quad (1.5.16)$$

$$= \frac{1}{n} \left[(b-a)^2 \int_a^b f^2(x) \frac{1}{b-a} dx - I^2 \right], \quad (1.5.17)$$

$$= \frac{1}{n} \left[(b-a) \int_a^b f^2(x) dx - I^2 \right]. \quad (1.5.18)$$

La méthode de Monte Carlo moyenne échantillonnale est appelée aussi en anglais méthode Crude Monte Carlo.

Algorithme

Algorithme de Monte Carlo moyenne échantillonnale

- (1) Générer une suite de n variables aléatoires $\left\{U_i\right\}_{i=1}^n$.
- (2) Calculer $X_i = a + U_i(b-a)$ pour $i = 1, \dots, n$.
- (3) Calculer $g(X_i)$ pour $i = 1, \dots, n$.
- (4) Calculer la moyenne échantillonnale par $\hat{\theta}_{Me} = (b-a) \frac{1}{n} \sum_{i=1}^n f(X_i)$.

1.5.3.5. Efficacité de la méthode de Monte Carlo

Supposons deux méthodes de Monte Carlo pour estimer l'intégrale (1.5.14). Soient $\hat{\theta}_1$ et $\hat{\theta}_2$ deux estimateurs obtenus par les deux méthodes tels que :

$$E(\hat{\theta}_1) = E(\hat{\theta}_2) = I.$$

On note t_1 et t_2 les temps nécessaires pour trouver $\hat{\theta}_1$ et $\hat{\theta}_2$ respectivement. Soit $\text{Var}(\hat{\theta}_1)$ la variance associée à la première méthode et $\text{Var}(\hat{\theta}_2)$ la variance associée à la deuxième méthode. Alors on dit que la première méthode est plus efficace que la deuxième méthode si

$$\epsilon = \frac{t_1 \text{Var}(\hat{\theta}_1)}{t_2 \text{Var}(\hat{\theta}_2)} < 1.$$

Comparons maintenant la méthode de Noir ou Blanc de Monte Carlo et la méthode de Monte Carlo moyenne en terme d'efficacité.

Proposition 1.5.1. *La méthode de moyenne échantionnale de Monte Carlo est plus efficace que la méthode de Noir ou Blanc de Monte Carlo lorsque les temps de calculs sont les mêmes car*

$$\text{Var}(\hat{\theta}_{Me}) \leq \text{Var}(\hat{\theta}_{NB}).$$

Démonstration

En soustrayant (1.5.18) de (1.5.10), on obtient

$$\text{Var}(\hat{\theta}_{NB}) - \text{Var}(\hat{\theta}_{Me}) = \frac{1}{n}(b-a) \left[cI - \int_a^b f^2(x)dx \right].$$

On rappelle que par (1.5.1)

$$f(x) \leq c,$$

donc

$$cI - \int_a^b f^2(x)dx \geq 0.$$

Ainsi

$$\text{Var}(\hat{\theta}_{NB}) - \text{Var}(\hat{\theta}_{Me}) \geq 0.$$

Chapitre 2

TECHNIQUES DE RÉDUCTION DE LA VARIANCE

2.1. INTRODUCTION

Dans cette partie, il est très important de rappeler que lors d'une estimation d'un paramètre, un bon estimateur est celui qui rend faible à la fois le biais au carré et la variance.

Plusieurs méthodes de Monte Carlo ont en commun le fait de ne pas introduire de biais dans l'estimation. Ainsi, lorsque le biais est nul, on peut se donner comme objectif de réduire au maximum la variance lors de l'estimation des paramètres.

Pour y arriver, plusieurs techniques ont été élaborées, entre autres l'échantillonnage pondéré, l'échantillonnage stratifié, les variables de contrôle et les variables antithétiques.

2.2. ÉCHANTILLONNAGE STRATIFIÉ

2.2.1. Introduction et principe général

Le principe général de cette méthode est de partitionner l'intervalle $(0, 1)$ en plusieurs sous-intervalles appelés strates, voir Hammersley & Handscomb (1964).

Considérons la partition suivante : $0 = \alpha_0 < \alpha_1 < \dots < \alpha_k = 1$ et appliquons la méthode de Monte Carlo élémentaire séparément sur chaque strate.

On cherche à estimer l'intégrale suivante :

$$\theta = \int_0^1 f(x)dx. \quad (2.2.1)$$

L'estimateur de θ est alors de la forme

$$\hat{\theta}_{Str} = \sum_{j=1}^k \sum_{i=1}^{n_j} \left((\alpha_j - \alpha_{j-1}) \frac{1}{n_j} f(\alpha_{j-1} + (\alpha_j - \alpha_{j-1})\xi_{ij}) \right) \quad (2.2.2)$$

où les ξ_{ij} sont des variables uniformément distribuées sur l'intervalle $[0, 1]$. Si on décide a priori de prendre un échantillon de taille n_j dans la j^e strate alors l'estimateur $\hat{\theta}_{Str}$ dans (2.2.2) est sans biais et sa variance est égale à

$$\text{Var}(\hat{\theta}_{Str}) = \sum_{j=1}^k \left(\frac{(\alpha_j - \alpha_{j-1})}{n_j} \int_{\alpha_{j-1}}^{\alpha_j} f^2(x)dx - \sum_{j=1}^k \left\{ \int_{\alpha_{j-1}}^{\alpha_j} f(x)dx \right\}^2 \right). \quad (2.2.3)$$

Cette variance peut être inférieure à $\text{Var}(\bar{f})$ avec $n = \sum_{j=1}^k n_j$, si la stratification a été suffisamment bien faite et le choix des n_j , $j = 1, \dots, k$ est adéquat.

Quand les limites de chaque strate sont fixées à l'avance, la meilleure façon de distribuer les points d'échantillon parmi les strates est de faire en sorte que n_j^2 soit proportionnel à :

$$\left[(\alpha_j - \alpha_{j-1}) \int_{\alpha_{j-1}}^{\alpha_j} f^2(x)dx - \left\{ \int_{\alpha_{j-1}}^{\alpha_j} f(x)dx \right\}^2 \right]. \quad (2.2.4)$$

Il y a différentes façons de choisir les α_j .

(1) La plus simple est de diviser l'intervalle en k parties égales,

$$\alpha_j = \frac{j}{k}.$$

(2) La meilleure façon est de choisir les α_j est plus compliquée. Intuitivement, on peut chercher à ce que la variance conditionnelle de f étant donnée qu'on se trouve dans une strate fixée soit constante d'une strate à l'autre.

Pour estimer l'erreur-type, on doit remplacer par une estimation dépendant de l'échantillon lui-même.

$$s_{\hat{\theta}_{str}}^2 = \sum_{j=1}^k \frac{(\alpha_j - \alpha_{j-1})^2}{n_j(n_j - 1)} \sum_{i=1}^{n_j} (f_{ij} - \bar{f}_j)^2, \quad (2.2.5)$$

$$\text{avec } f_{ij} = f(\alpha_{j-1} + (\alpha_j - \alpha_{j-1})\xi_{ij}), \quad (2.2.6)$$

$$\text{et } \bar{f}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} f_{ij}. \quad (2.2.7)$$

Soit l'intégrale suivante :

$$I = \int_D g(x)f(x)dx$$

où f est une fonction densité et g^2 est sommable sur D .

Supposons qu'on stratifie notre région en m strates D_i , $i = 1, \dots, m$, c'est-à-dire, $D = \bigcup_{i=1}^m D_i$, $D_k \cap D_j = \emptyset$, $k \neq j$
et qu'on veuille calculer

$$I_i = \int_{D_i} g(x)f(x)dx.$$

Définissons

$$P_i = \int_{D_i} f(x)dx.$$

Il est évident que $\sum_{i=1}^m P_i = 1$ et

$$I = \int_D g(x)f(x)dx = \sum_{i=1}^m \int_{D_i} g(x)f(x)dx = \sum_{i=1}^m I_i.$$

Soit

$$g_i(x) = \begin{cases} g(x) & \text{si } x \in D_i, \\ 0 & \text{sinon.} \end{cases}$$

On réécrit l'intégrale I_i comme étant

$$\begin{aligned}
 I_i &= \int_{D_i} P_i g(x) \frac{f(x)}{P_i} dx, \\
 &= P_i \int_{D_i} g_i(x) \frac{f(x)}{P_i} dx, \\
 &= P_i E[g_i(X)], \\
 &= P_i E[g(X) | X \in D_i].
 \end{aligned}$$

Puisque $E[g(X) | X \in D_i]$ peut être exprimée comme l'espérance d'une variable aléatoire X_i de densité f_i avec,

$$f_i(x) = \frac{f(x)}{P_i} \quad x \in D_i,$$

l'intégrale I_i est estimée par

$$\tau_i = \frac{P_i}{N_i} \sum_{k=1}^{N_i} g(X_{ik}), \quad i = 1, \dots, m$$

où les X_{ik} , $k = 1, \dots, N_i$, sont des variables aléatoires indépendantes de même loi que X_i , $i = 1, \dots, m$.

Ainsi, l'intégrale I est estimée par

$$\hat{\theta}_{Str} = \sum_{i=1}^m \tau_i = \sum_{i=1}^m \frac{P_i}{N_i} \sum_{k=1}^{N_i} g(X_{ki}).$$

On a donc

$$\text{Var}(\hat{\theta}_{Str}) = \sum_{i=1}^m \frac{P_i^2}{N_i} \text{Var}(g(X_i)) = \sum_{i=1}^m \frac{P_i^2 \sigma_i^2}{N_i},$$

où

$$\sigma_i^2 = \text{Var}(g(X_i)) = \frac{1}{P_i} \int_{D_i} g^2(x) f(x) dx - \frac{I_i^2}{P_i^2}. \quad (2.2.8)$$

Théorème 2.2.1. *Si la taille de chaque strate est proportionnelle à P_i , $N_i = NP_i$, alors*

$$\text{Var}(\hat{\theta}_{Str}) \leq \text{Var}(\hat{\theta}_{Me}).$$

Démonstration

Substituons $N_i = NP_i$ dans

$$\text{Var}(\hat{\theta}_{Str}) = \sum_{i=1}^m \frac{P_i^2}{N_i} \text{Var}(g(X_i)).$$

Nous obtenons

$$\text{Var}(\hat{\theta}_{Str}) = \frac{1}{N} \sum_{i=1}^m P_i \text{Var}(g(X_i)).$$

Par l'inégalité de Cauchy-Schwarz nous avons

$$\begin{aligned} I^2 &= \left(\sum_{i=1}^m I_i \right)^2, \\ &= \left[\sum_{i=1}^m \frac{I_i}{\sqrt{P_i}} \sqrt{P_i} \right]^2, \\ &\leq \sum_{i=1}^m \frac{I_i^2}{P_i} \sum_{i=1}^m P_i = \sum_{i=1}^m \frac{I_i^2}{P_i}. \end{aligned}$$

Multiplions par P_i dans

$$\sigma_i^2 = \text{Var}(g(X_i)) = \frac{1}{P_i} \int_{D_i} g^2(x) f(x) dx - \frac{I_i^2}{P_i^2},$$

et faisons la somme sur i qui varie de 1 à m . Nous obtenons

$$\sum_{i=1}^m \text{Var}(P_i g(X_i)) = \int_D g^2(x) f(x) dx - \sum_{i=1}^m \frac{I_i^2}{P_i}.$$

On peut écrire

$$\sum_{i=1}^m \text{Var}(P_i g(X_i)) \leq \int_D g^2(x) f(x) dx - I^2 = N \text{Var}(\hat{\theta}_{Me}).$$

Dans le cas discret, on va considérer tout au long de ce travail deux variables X et Y dont Y sera la variable d'intérêt et X la variable explicative. L'objectif de la simulation est d'estimer $E(Y) = \theta$.

Pour l'échantillonnage stratifié, la connaissance de la distribution marginale de X est indispensable pour s'assurer que les proportions échantillonnales des n différentes valeurs de X sont exactement égales à leurs probabilités de réalisations.

Si l'on considère que n_i est le nombre de fois que $X = i$ dans l'échantillon alors

$$n_i = nP(X = i).$$

Soit Y_{ij} , $j = 1, 2, \dots, n_i$, les n_i valeurs de Y observées quand $X = i$ et on définit $\bar{Y}_i = \sum_{j=1}^{n_i} Y_{ij} / n_i$ la moyenne des Y_{ij} , $j = 1, \dots, n_i$.

Donc

$$\hat{\theta}_{Str} = \sum_i \bar{Y}_i P(X = i).$$

Ce qui correspond à estimer $E(Y|X = i)$ par \bar{Y}_i dans

$$E(Y) = \sum_i E(Y|X = i)P(X = i). \quad (2.2.9)$$

Cet estimateur est sans biais et la variance de $\hat{\theta}_{Str}$ est

$$\begin{aligned}
 \text{Var}(\hat{\theta}_{Str}) &= \sum_i P^2(X = i) \text{Var}(\bar{Y}_i), \\
 &= \frac{1}{n} \sum_i \text{Var}(Y|X = i) P(X = i), \\
 &= \frac{1}{n} E(\text{Var}(Y|X)), \\
 &= \frac{1}{n} \left[\text{Var}(Y) - \text{Var}(E(Y|X)) \right].
 \end{aligned}$$

Exemple 2.2.1.

On imagine un projet où on doit faire travailler des hommes. Soit Y le nombre de personnes affectées pour le travail. Supposons que l'on veut associer le nombre d'hommes, Y , affectés chaque jour tout au long de l'étude du projet. L'affectation se fait quotidiennement et chaque jour de recrutement est indépendant de l'autre. Le nombre de gens, X , disponible pour l'affectation varie aussi quotidiennement.

La variable X est aléatoire avec probabilité de distribution $P(X = 1) = 0,4$, $P(X = 2) = 0,45$ et $P(X = 3) = 0,25$. La variable Y est déterminée selon X de la façon suivante :

Si c'est un seul employé qui est disponible, il doit travailler sur le projet avec probabilité de 0,4.

Si ce sont deux employés qui sont disponibles ($X = 2$) alors le premier sera affecté avec probabilité de 0,71 et le second sera seulement choisi si le premier l'est et en plus il sera affecté avec probabilité de 0,72.

Si les trois employés sont disponibles, l'un des trois sera affecté automatiquement sur le projet, l'autre sera refusé automatiquement et le dernier sera affecté avec probabilité de 0,8.

Soit donc le tableau suivant qui définit les probabilités conjointes $P(X = i, Y = j)$ de nos deux variables aléatoires X et Y ,

		X			$P(Y = j)$
		1	2	3	
Y	2	0,0	0,09	0,2	0,29
	1	0,12	0,23	0,05	0,40
	0	0,18	0,13	0,0	0,31
$P(X = i)$		0,3	0,45	0,25	

TABLE 2.1. Tableau des probabilités conjointes de (X, Y)

Par des calculs simples on a les résultats suivants :

$$E(X) = 1,95, \quad E(Y) = 0,98,$$

$$\text{Var}(X) = 0,5475, \quad \text{Var}(Y) = 0,5996, \quad \text{Cov}(X, Y) = 0,379.$$

$$E(\text{Var}(Y|X)) = 0,3285, \quad \text{Var}(E(Y|X)) = 0,2711.$$

TABLE 2.2. Tableau des espérances et variances de $(Y|X = i)$

i	1	2	3
$E(Y X=i)$	0,4	0,91	1,8
$\text{Var}(Y X=i)$	0,24	0,48	0,16

Par exemple, si $n = 100$ alors on fait trois strates de tailles 30, 45 et 25. Par la formule de l'échantillonnage stratifié on obtient :

$$\text{Var}(\hat{\theta}_{Str}) = 0,003285 \text{ car } E(\text{Var}(Y|X)) = 0,3285.$$

2.3. ÉCHANTILLONNAGE PONDÉRÉ

2.3.1. Principe général

Cette méthode est aussi appelée échantillonnage selon l'importance du fait de la présence de la fonction de poids dans l'expression de l'estimateur. La méthode de Monte Carlo moyenne échantillonnale exposée à la fin du chapitre 1 en est un cas particulier.

L'idée fondamentale de cette technique est de concentrer la distribution des points de l'échantillon dans les parties de la région qui sont plus «importantes» au lieu de les disperser au hasard.

Nous voulons évaluer l'intégrale

$$I = \int_{\mathcal{X}} h(x) f(x) dx. \quad (2.3.1)$$

On appelle échantillonnage pondéré toute méthode proposant d'approcher I à partir de l'échantillon X_1, X_2, \dots, X_n généré selon une loi de densité g par l'approximation

$$\widehat{I}_n = \frac{1}{n} \sum_{j=1}^n \frac{f(x_j)}{g(x_j)} h(x_j). \quad (2.3.2)$$

Cette expression (2.3.2) provient d'une représentation alternative de I à savoir

$$J = \int_{\mathcal{X}} h(x) \frac{f(x)}{g(x)} g(x) dx. \quad (2.3.3)$$

L'estimateur \widehat{I}_n converge vers I . La loi de densité g est dite «loi instrumentale» ou «loi d'importance». Elle peut être sélectionnée parmi les lois faciles à simuler.

L'estimateur de J ne possède une variance finie que si l'espérance

$$E_f \left[h^2(X) \frac{f(X)}{g(X)} \right] = \int_{\mathcal{X}} h^2(x) \frac{f^2(x)}{g(x)} dx \quad \text{est finie.} \quad (2.3.4)$$

De plus

$$\begin{aligned} \text{Var}(\widehat{I}_n) &= \frac{1}{n} \text{Var} \left(h(X) \frac{f(X)}{g(X)} \right) \quad \text{où } X \text{ est de densité } g \\ &= \frac{1}{n} \int_{\mathcal{X}} \left(h(x) \frac{f(x)}{g(x)} - I \right)^2 g(x) dx. \end{aligned}$$

2.3.2. Choix optimal de la fonction d'importance

Il importe ici de choisir la fonction g qui facilitera la réduction de la variance.

Théorème 2.3.1. *Le choix de g minimisant la variance de l'estimateur (2.3.1) est*

$$g^*(x) = \frac{|h(x)|f(x)}{\int_{\mathcal{X}} |h(z)|f(z)dz}.$$

Démonstration

$$\begin{aligned} E_f \left[h^2(X) \frac{f(X)}{g^*(X)} \right] &= \tilde{I} E_f[|h(X)|^2/|h(X)|] \\ &= \tilde{I} E_f[|h(X)|] = \tilde{I}^2, \end{aligned}$$

$$\text{avec } \tilde{I} = \int_{\mathcal{X}} |h(x)|f(x)dx.$$

On a

$$\begin{aligned} \left(\int_{\mathcal{X}} |h(x)|f(x)dx \right)^2 &= \left(\int_{\mathcal{X}} |h(x)| \frac{f(x)}{g(x)} g(x)dx \right)^2, \\ &\leq \left(\int_{\mathcal{X}} h^2(x) \frac{f^2(x)}{g(x)}dx \right) \int_{\mathcal{X}} g(x)dx \end{aligned}$$

pour toute densité g , par l'inégalité de Cauchy-Schwarz.

Corollaire 2.3.1. *Si $h(x)=1$ et $f(x)>0$, alors la densité optimale est*

$$g(x) = \frac{f(x)}{\int f(x)dx}.$$

Cette optimalité n'est pas très utile car lorsque $h(x) > 0$, le choix optimal de $g^*(x) = h(x)f(x)/I$, nous oblige à connaître l'intégrale I , objet de l'étude. Éliminer ce problème en considérant l'alternative à (2.3.1)

$$\frac{\sum_{j=1}^n h(x_j)f(x_j)/g(x_j)}{\sum_{j=1}^n f(x_j)/g(x_j)} \quad (2.3.5)$$

autorise à utiliser des densités f et g connues à un facteur multiplicatif près mais, bien que (2.3.2) converge également vers I en vertu de la loi des grands nombres, l'optimalité ne se transmet pas à (2.3.2). De plus (2.3.2) est biaisé pour n fini et lorsque $E_f[f(x)/g(x)]$ est infini, cet estimateur a un comportement potentiellement très instable. Casella & Robert(1996) ont cependant montré de meilleures performances de cet estimateur pondéré dans certaines situations.

Mais du point de vue pratique, on cherche à utiliser des densités g telles que $|h|f/g$ soit pratiquement constant et qui vérifient les contraintes de finitude sur la variance.

2.3.3. Algorithme

L'algorithme de simulation de Monte Carlo pour la méthode d'échantillonnage pondéré est :

- (1) Générer n nombres aléatoires $\{X_i\}_{i=1}^n$ selon la densité g .
- (2) Calculer $\gamma(X_i) = h(X_i)\frac{f(X_i)}{g(X_i)}$ et $d(X_i) = f(X_i)/g(X_i)$ pour $i = 1, \dots, n$.
- (3) Estimer I par $\sum_{i=1}^n \gamma(X_i) / \sum_{i=1}^n d(X_i)$.

Dans le cas discret, nous allons utiliser la même expression que précédemment

$$\hat{\theta}_{imp} = \sum_i \bar{Y}_i P(X = i) \quad (2.3.6)$$

qui correspond à estimer θ par la formule (2.2.9) avec $E(Y|X = i)$ remplacé par \bar{Y}_i .

Donc (2.3.6) peut s'écrire comme :

$$\hat{\theta}_{imp} = \sum_i (\bar{Y}_i r_i) P(Z = i)$$

où

$$r_i = \frac{P(X = i)}{P(Z = i)}.$$

Ici, Z est une variable aléatoire avec

$$P(Z = i) = \frac{n_i}{n} \text{ pour } i = 1, \dots, n.$$

De (2.3.6) on a

$$\text{Var}(\hat{\theta}_{imp}) = \sum_i \frac{P^2(X = i)}{n_i} \text{Var}(Y|X = i).$$

Pour être sûr de minimiser la variance, il faut choisir n_i tel que

$$n_i = kP(X = i)\sqrt{(\text{Var}(Y|X = i))} \quad (2.3.7)$$

qui satisfait la contrainte $n = \sum_i n_i$ où k est la constante de proportionnalité.

Ce choix optimal est appelé échantillonnage élaboré parce qu'il affecte plus de poids aux valeurs de X pour lesquelles $P(X = i)$ est plus élevée et pour lesquelles aussi la variance conditionnelle de Y est plus grande.

Exemple 2.3.1.

En considérant le même exemple (2.2.1) on est capable de trouver que $n_0 = 17$, $n_1 = 72$ et $n_2 = 11$ par la formule (2.3.7) et en substituant les valeurs de $P(X = i)$ et $\text{Var}(Y|X = i)$ dans la formule (2.3.7) qu'on trouve respectivement dans les tableaux (2.1) et (2.2) on trouve

$$\text{Var}(\hat{\theta}_{imp}) = 0,0031219.$$

2.4. VARIABLES DE CONTRÔLE

2.4.1. Idée générale

L'idée principale de cette technique de réduction de la variance lors d'une simulation consiste à produire par simulation une variable auxiliaire qui est en corrélation positive avec la variable d'intérêt.

Soit X la variable aléatoire dont on veut estimer la moyenne $\mu = E(X)$, Y la variable d'intérêt obtenue par simulation et $\theta = E(Y)$ sa moyenne. Ici, Y est appelée variable de contrôle.

Construisons l'estimateur de μ de la façon suivante :

$$X_c = X - a(Y - \theta) \quad \text{avec} \quad a \in \mathbb{R}.$$

Ici

$$\begin{aligned} E(X_c) &= E(X - a(Y - \theta)), \\ &= E(X) - a(E(Y - \theta)), \\ &= E(X) - aE(Y) + a\theta, \\ &= E(X) - a\theta + a\theta, \\ &= E(X) = \mu. \end{aligned}$$

Donc X_c est un estimateur sans biais de μ , $\forall a \in \mathbb{R}$.

On obtient

$$\begin{aligned} \text{Var}(X_c) &= \text{Var}(X - a(Y - \theta)), \\ &= \text{Var}(X) + \text{Var}(a(Y - \theta)) - 2\text{Cov}(X, a(Y - \theta)), \\ &= \text{Var}(X) + a^2\text{Var}(Y - \theta) - 2a\text{Cov}(X, (Y - \theta)), \\ &= \text{Var}(X) + a^2\text{Var}(Y) - 2a\text{Cov}(X, Y). \end{aligned}$$

Donc

$$\text{Var}(X_c) < \text{Var}(X) \quad \text{si} \quad a^2 \text{Var}(Y) - 2a \text{Cov}(X, Y) < 0,$$

c'est-à-dire que

$$0 < a < \frac{2\text{Cov}(X, Y)}{\text{Var}(Y)}.$$

Alors $\text{Var}(X_c)$ sera minimisée par la valeur optimale a^* de a

$$a^* = \frac{\text{Cov}(X, Y)}{\text{Var}(Y)}.$$

Ainsi la valeur optimale de $\text{Var}(X_c)$ est $\text{Var}(X_c^*)$ avec

$$\begin{aligned} \text{Var}(X_c^*) &= \text{Var}(X) + \left[\frac{\text{Cov}(X, Y)}{\text{Var}(Y)} \right]^2 \text{Var}(Y) - 2 \left[\frac{\text{Cov}(X, Y)}{\text{Var}(Y)} \right] \text{Cov}(X, Y), \\ &= \text{Var}(X) - \left[\frac{\text{Cov}(X, Y)^2}{\text{Var}(Y)} \right], \\ &= \left[1 - \rho_{XY}^2 \right] \text{Var}(X). \end{aligned}$$

Ici

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

et ρ_{XY} est appelé le coefficient de corrélation entre X et Y , $-1 \leq \rho_{XY} \leq 1$.

Dans le meilleur des cas possibles, si le coefficient de corrélation est égal à -1 ou 1, la variance est réduite à zéro.

Dans le pire des cas, quand il n'y a pas de corrélation, $\rho_{XY} = 0$ et la variance reste inchangée, c'est-à-dire $\text{Var}(X_c^*) = \text{Var}(X)$.

Dépendemment de la source ou de la nature de la variable de contrôle Y , on peut connaître ou ne pas connaître $\text{Var}(Y)$ mais $\text{Cov}(X, Y)$ est le plus souvent inconnu.

Si $\text{Cov}(X, Y)$ est inconnu alors la valeur exacte de a^* est aussi inconnue. Plusieurs méthodes de simulation ont été proposées pour estimer la valeur de a^* . L'une des méthodes consiste à se donner un échantillon de taille n avec X_1, X_2, \dots, X_n des variables *iid* et un vecteur de contrôle Y_1, Y_2, \dots, Y_n , de n variables *iid*.

Soient $\bar{X}_{(n)} = \sum_{i=1}^n X_i/n$, $\bar{Y}_{(n)} = \sum_{i=1}^n Y_i/n$, les moyennes échantionnales respectivement des X_i et Y_i et $s_{Y(n)}^2$ l'estimateur sans biais de $\text{Var}(Y)$.

L'estimateur de la covariance entre X et Y est :

$$\widehat{\text{Cov}(X, Y)}_{(n)} = \frac{\sum_{i=1}^n [X_i - \bar{X}_{(n)}][Y_i - \bar{Y}_{(n)}]}{n - 1}.$$

L'estimateur de a^* est alors

$$\widehat{a^*} = \frac{\widehat{\text{Cov}(X, Y)}_{(n)}}{s_{Y(n)}^2}.$$

Ainsi l'estimateur de μ est

$$\widehat{X}_{c(n)} = \bar{X}_{(n)} - \widehat{a^*} [\bar{Y}_{(n)} - \theta].$$

L'idée généralisée considère m variables de contrôle Y^1, Y^2, \dots, Y^m avec leurs moyennes $\theta^1, \theta^2, \dots, \theta^m$.

L'estimateur sera donc :

$$X_c = X - \sum_{j=1}^m a_j (Y^j - \theta^j)$$

où les a_j sont des réels à déterminer ou à estimer.

on aura :

$$\begin{aligned}\text{Var}(X_c) &= \text{Var}(X) + \sum_{j=1}^m a_j^2 \text{Var}(Y^j) \\ &\quad - 2 \sum_{j=1}^m a_j (\text{Cov}(X, Y^j)) + 2 \sum_{i=2}^m \sum_{j=1}^{i-1} a_i a_j \text{Cov}(Y^i, Y^j).\end{aligned}$$

Les a_i et a_j sont obtenus de la même façon que les estimateurs des coefficients d'un modèle de régression linéaire.

Exemple 2.4.1.

Pour le même exemple (2.2.1), nous avons trouvé

$$\text{Var}(X) = 0,5475, \quad \text{Cov}(X, Y) = 0,379 \text{ et } \text{Var}(Y) = 0,5996.$$

On trouve le choix optimal de a

$$a = 0,379/0,5475 = 0,692$$

et

$$\text{Var}(\hat{\theta}_{Cont}) = 0,3376/n.$$

2.5. VARIABLES ANTITHÉTIQUES

Historiquement le nom des variables antithétiques a été introduit par Hamersley & Mauldon (1956). Contrairement à la méthode des variables de contrôle, la méthode des variables antithétiques introduit une corrélation négative entre les variables.

2.5.1. Idée générale

Soient X et Y deux variables aléatoires avec fonctions de répartition F_X et F_Y respectivement, avec F_X et F_Y connues. On cherche un estimateur de l'espérance de $X + Y$ avec une variance minimale.

Or

$$\text{Var}\left(\frac{1}{2}(X + Y)\right) = \frac{1}{4}\text{Var}(X) + \frac{1}{4}\text{Var}(Y) + \frac{1}{2}\text{Cov}(X, Y).$$

Donc minimiser cette variance revient à minimiser la covariance. Pour être capable de réaliser la réduction de cette variance il faut rendre la covariance négative, ainsi la variance sera plus petite que dans le cas où les variables sont indépendantes (covariance nulle).

Supposons que X et Y sont des variables générées par la méthode de transformation inverse

$$X = F_X^{-1}(U_1) = \inf \left\{ x : F_X(x) \geq U_1 \right\},$$

$$Y = F_Y^{-1}(U_2) = \inf \left\{ y : F_Y(y) \geq U_2 \right\},$$

où U_1 et U_2 sont de lois uniformes sur $(0, 1)$.

La technique des variables antithétiques consiste à prendre $U_2 = 1 - U_1$ ce qui nous amène au théorème suivant.

Théorème 2.5.1. *Soient U_1, U_2, \dots, U_d des variables aléatoires iid uniformes sur $(0, 1)$, soit $k : [0, 1]^d \longrightarrow \mathbb{R}$ une fonction croissante. Soient $\alpha = k(U_1, U_2, \dots, U_d)$ et $\beta = k(1 - U_1, 1 - U_2, \dots, 1 - U_d)$. On obtient alors,*

(1) α et β ont même distribution,

(2) $\text{Cov}(\alpha, \beta) \leq 0$.

Considérons l'intégrale suivante :

$$I = \int_0^1 g(x) dx.$$

Donc

$$I = \frac{1}{2} \int_0^1 \left\{ g(x) + g(1 - x) \right\} dx.$$

Alors l'estimateur de I est

$$\hat{\theta} = \frac{1}{2}(X + Y) = \frac{1}{2} \left\{ g(U) + g(1 - U) \right\}.$$

L'estimateur $\hat{\theta}$ est sans biais de I car $X = g(U)$ et $Y = g(1 - U)$ sont des estimateurs sans biais de I .

En général pour estimer I on considère un échantillon de taille n d'une distribution uniforme et on calcule :

$$\hat{\theta}_{Ant} = \frac{1}{2n} \sum_{i=1}^n [g(U_i) + g(1 - U_i)].$$

2.5.2. Variables antithétiques pour des fonctions monotones

Proposition 2.5.1. *Si g est une fonction monotone non croissante (non décroissante) et dérivable alors*

$$\text{Var}(\hat{\theta}_{Ant}) \leq \frac{1}{2} \text{Var}(\hat{\theta}_{Me}).$$

Démonstration

Supposons sans perte de généralité que $n = 1$. On obtient

$$\begin{aligned} \text{Var}(\hat{\theta}_{Ant}) &= \frac{1}{4} \int_0^1 g^2(x) dx + \frac{1}{4} \int_0^1 g^2(1-x) dx + \frac{1}{2} \int_0^1 g(x)g(1-x) dx - I^2, \\ &= \frac{1}{2} \int_0^1 g^2(x) dx + \frac{1}{2} \int_0^1 g(x)g(1-x) dx - I^2. \end{aligned}$$

Or d'après (1.5.15) on sait que

$$\text{Var}(\hat{\theta}_{Me}) = \int_0^1 g^2(x) dx - I^2.$$

Donc

$$2\text{Var}(\hat{\theta}_{Ant}) - \text{Var}(\hat{\theta}_{Me}) = \int_0^1 g(x)g(1-x) dx - I^2.$$

On remarque que la proposition sera démontrée si on parvient à montrer que

$$\int_0^1 g(x)g(1-x) dx \leq I^2.$$

Introduisons une autre fonction auxiliaire

$$\phi(x) = \int_0^x g(1-t)dt - xI,$$

telle que

$$\phi(0) = \phi(1) = 0.$$

La première dérivée de ϕ donne

$$\phi'(x) = g(1-x) - I,$$

donc ϕ' est décroissante ce qui implique que ϕ est concave et par le fait même ϕ est positive. Comme ϕ et g' sont positives on obtient :

$$\int_0^1 \phi(x)g'(x) \geq 0.$$

En intégrant par partie on trouve que

$$\int_0^1 \phi'(x)g(x) \leq 0.$$

Si on substitue

$$\phi'(x) = g(1-x) - I,$$

dans

$$\int_0^1 \phi'(x)g(x) \leq 0,$$

alors on obtient

$$\int_0^1 g(x)g(1-x)dx \leq I^2.$$

Dans la suite nous allons traiter un cas beaucoup plus général où il faut estimer

$$\theta = E(g(X) + h(Y)) \quad \text{avec variance minimale,} \quad (2.5.1)$$

avec X et Y des vecteurs aléatoires de dimension n , g et h des fonctions réelles mesurables. On fixe les distributions marginales de X et de Y à F_X et F_Y et on pose Ξ_n comme étant l'espace de toutes les fonctions de répartition qu'on peut mettre sur (X, Y) de sorte que la fonction de répartition marginale de X soit F_X et celle de Y soit F_Y . Il faut trouver une fonction de répartition $F \in \Xi_n$ qui minimise $\text{Var}(g(X) + h(Y))$.

Gal, Rubinstein, & Ziv (1984) ont pu démontrer que si g et h sont des fonctions monotones le même sens alors

$$\begin{aligned} \min_{f \in \Xi_n} \text{Var}_{F_{XY}} \{g(X) + h(Y)\} &= \text{Var}\{g(F_X^{-1}(U_1)) + h(F_Y^{-1}(U_2))\}, \quad (2.5.2) \\ &= \text{Var}\{g^*(U_1) + h^*(U_2)\}, \\ &= \text{Var}\{g^*(U) + h^*(1 - U)\}, \end{aligned}$$

où les distributions F_X et F_Y sont connues sur \mathbb{R}^n ,

$$F_X = \Pi_{i=1}^n F_{X_i}(x_i) \quad \text{et} \quad F_Y = \Pi_{i=1}^n F_{Y_i}(y_i),$$

$$\text{avec } X = F_X^{-1}(U_1) = \{F_X^{-1}(U_1^1), \dots, F_X^{-1}(U_1^n)\}, \quad (2.5.3)$$

$$\text{et } Y = F_Y^{-1}(1 - U_1) = \{F_Y^{-1}(1 - U_1^1), \dots, F_Y^{-1}(1 - U_1^n)\}, \quad (2.5.4)$$

donc

$$g^*(U_1) = g(F_X^{-1}(U_1)) \quad \text{et} \quad h^*(U_2) = h(F_Y^{-1}(U_2)).$$

Ceci signifie que l'utilisation du nombre aléatoire antithétique, $U_1 = U$ et $U_2 = 1 - U_1 = 1 - U$, est optimal pour le problème (2.5.2) où U est un vecteur d'uniformes $(0, 1)$ indépendantes.

Mitchell (1973) a suggéré d'utiliser le nombre aléatoire antithétique ($U_2 = 1 - U_1$) pour estimer $E(g^*(U_1) + h^*(U_2))$, $U_1, U_2 \in \mathbb{R}^n$; il a pu démontrer que si $g^*(u_1)$ et $h^*(u_2)$, $u_1, u_2 \in \mathbb{R}^n$ sont des fonctions monotones dans les mêmes directions à chaque argument du vecteur u , alors

$$\text{Cov}(g^*(U), h^*(1 - U)) \leq 0, \quad (2.5.5)$$

où $U = (U^{(1)}, \dots, U^{(n)})$ est un vecteur de variables aléatoires indépendantes, chacune uniformément distribuée sur $[0, 1]$.

Ainsi pour l'estimation de (2.5.1) la réduction de la variance est obtenue en utilisant le vecteur aléatoire antithétique.

De (2.5.5) on a

$$\text{Var}[g^*(U_1) + h^*(U_2)] \geq \text{Var}[g^*(U) + h^*(1 - U)]. \quad (2.5.6)$$

Variables communes

Contrairement à la méthode de variables antithétiques, la méthode de variables communes consiste à estimer

$$\theta = E(X - Y) \quad \text{avec variance minimale,} \quad (2.5.7)$$

donc cela revient à minimiser

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y).$$

Dans ce cas on suppose que les distributions marginales de X et Y sont connues. Donc il est évident que minimiser $\text{Var}(X - Y)$ revient à maximiser la covariance de X et Y .

Considérons maintenant l'estimation de

$$\theta = E(g(X) - h(Y)) \quad \text{avec variance minimale} \quad (2.5.8)$$

où X et Y sont des vecteurs aléatoires de dimension n , g et h sont des fonctions réelles mesurables, et les distributions marginales F_X et F_Y sont connues. Mathématiquement le problème se pose comme ceci :

$$\text{minimiser } \text{Var}_F(g(X) - h(Y)), \quad F \in \Xi_n$$

où les distributions F_X et F_Y sur \mathbb{R}^n sont connues et sont données par :

$$F_X = \Pi_{i=1}^n F_{X_i}(x_i) \quad \text{et} \quad F_Y = \Pi_{i=1}^n F_{Y_i}(y_i)$$

$$\text{avec } X = F_X^{-1}(U_1) = \{F_X^{-1}(U_1^1), \dots, F_X^{-1}(U_1^n)\}, \quad (2.5.9)$$

$$\text{et } Y = F_Y^{-1}(U_2) = \{F_Y^{-1}(U_2^1), \dots, F_Y^{-1}(U_2^n)\}. \quad (2.5.10)$$

En définissant

$$g^*(U_1) = g(F_X^{-1}(U_1)) \quad \text{et} \quad h^*(U_2) = h(F_Y^{-1}(U_2)),$$

et si g et h sont monotones dans les mêmes directions alors

$$\min_{F \in \Xi_n} \text{Var}_F\{g(X) - h(Y)\} = \text{Var}\{g(F_X^{-1}(U_1)) - h(F_Y^{-1}(U_2))\}, \quad (2.5.11)$$

$$= \text{Var}\{g^*(U_1) - h^*(U_2)\}, \quad (2.5.12)$$

$$= \text{Var}\{g^*(U) - h^*(U)\}.$$

Ceci signifie que l'utilisation du vecteur aléatoire commun, $U_1 = U$ et $U_2 = U_1 = U$, est optimale pour le problème (2.5.12).

Donc Mitchel (1973) a suggéré d'utiliser le nombre aléatoire commun pour estimer $E(g^*(U_1) - h^*(U_2))$, $U_1, U_2 \in \mathbb{R}^n$. Il a pu démontrer que si $g^*(u_1)$ et $h^*(u_2)$, $u_1, u_2 \in \mathbb{R}^n$ sont des fonctions monotones dans la même direction à chaque argument du vecteur u , alors

$$\text{Cov}(g^*(U), h^*(U)) \geq 0, \quad (2.5.13)$$

où $U = (U^{(1)}, \dots, U^{(n)})$ est un vecteur de variables aléatoires indépendantes, chacune est distribuée uniformément sur $[0, 1]$.

De (2.5.13) on a

$$\text{Var}[g^*(U_1) - h^*(U_2)] \geq \text{Var}[g^*(U) - h^*(U)]. \quad (2.5.14)$$

Résultats théoriques

On peut réécrire le problème (2.5.2) comme ceci

$$\text{Minimiser } \text{Var}\{g^*(U_1) + h^*(U_2)\} \quad (2.5.15)$$

où U_1 et U_2 sont des vecteurs de variables aléatoires indépendantes, chacune est uniformément distribuée sur $[0, 1]$ et on définit

$$g(x) = g\{F_{X_1}^{-1}(u_1^{(1)}), \dots, F_{X_n}^{-1}(u_1^{(n)})\} \triangleq g^*(u_1)$$

$$h(y) = g\{F_{Y_1}^{-1}(u_2^{(1)}), \dots, F_{Y_n}^{-1}(u_2^{(n)})\} \triangleq h^*(u_2).$$

On distingue trois cas du problème. (2.5.15)

2.5.2.1. *Les fonctions g et h sont monotones dans chaque composante de x et y , $x, y \in \mathbb{R}^n$*

Dans cette situation $g^*(u_1)$ et $h^*(u_2)$ sont aussi des fonctions monotones dans chaque composante de u_1 et u_2 , respectivement, ainsi toutes les fonctions $F_{X_i}^{-1}$ et $F_{Y_i}^{-1}$, $i = 1, \dots, n$, sont monotones et non décroissantes.

Donc dans l'ensemble de toutes les fonctions de distribution $F_{U_1 U_2}$ sur \mathbb{R}^{2n} , où la dépendance est permise seulement entre les composantes de même indice de U_1 et U_2 alors le vecteur aléatoire antithétique avec $U_2 = 1 - U_1$, est optimal pour le problème (2.5.2), où 1 est un vecteur unité, si f et g sont des fonctions monotones dans la même direction. Donc le vecteur aléatoire $\{F_X^{-1}(U), F_Y^{-1}(1 - U)\}$ a une fonction de répartition avec corrélation minimum ce qui résout donc le problème (2.5.2).

2.5.2.2. *Les fonctions g et h sont monotones et croissantes pour certaines composantes et monotones décroissantes pour d'autres composantes*

Supposons que $g(u_1)$ et $h(u_2)$ sont monotones croissantes sur certaines composantes de u_1 et u_2 monotones décroissantes pour d'autres composantes, on traite cette situation en définissant le vecteur

$$\bar{u}_2^{(k)} = \begin{cases} u_1^{(k)}, & \text{si } g \text{ et } h \text{ sont monotones} \\ & \text{dans la direction opposée dans la } k^e \text{ composante.} \\ 1 - u_1^{(k)}, & \text{si } g \text{ et } h \text{ sont monotones} \\ & \text{dans la même direction dans la } k^e \text{ composante.} \end{cases} \quad (2.5.16)$$

Le vecteur $\bar{u}_2^{(k)}$ est appelé vecteur des variables aléatoires antithétique-commun.

Exemple 2.5.1.

Soit $g(u_1^{(1)}, u_1^{(2)})$ une fonction monotone croissante sur $u_1^{(1)}$ et monotone décroissante dans $u_1^{(2)}$ et $h(u_2^{(1)}, u_2^{(2)})$ une fonction monotone croissante sur tous les deux $u_2^{(1)}, u_2^{(2)}$. Alors

$$\bar{u}_2 = (\bar{u}_2^{(1)}, \bar{u}_2^{(2)}) = (u_1^{(1)}, 1 - u_1^{(2)})$$

est un vecteur aléatoire antithétique-commun qui maximise la covariance.

Soient

$$\begin{aligned} g(u_1^{(1)}, u_2^{(1)}) &= u_1^{(1)} + e^{-u_2^{(1)}} \\ \text{et } h(u_2^{(1)}, u_2^{(2)}) &= 5u_1^{(2)}(u_2^{(2)})^2 \end{aligned}$$

et on suppose que la dépendance est permise seulement entre les composantes souhaitées. Par simple calcul nous avons

$$E(g(U_1, U_2)) = E(U_1 + e^{-U_2}) = 1,132$$

$$E(h(U_1, U_2)) = E(5U_1U_2^2) = 0,833$$

$$E(g^2(U_1, U_2)) = 1,398, \quad E(h^2(U_1, U_2)) = 1,667$$

$$\text{Var}(g(U_1, U_2)) = E(g^2(U_1, U_2)) - (E(g(U_1, U_2)))^2 = 1,398 - (1,132)^2 = 0,116$$

$$\text{Var}(h(U_1, U_2)) = E(h^2(U_1, U_2)) - (E(h(U_1, U_2)))^2 = 0,972.$$

Avec des variables indépendantes on a

$$\begin{aligned} \text{Var}(g(U_1, U_2) + h(U_1, U_2)) &= \text{Var}(g(U_1, U_2) - h(U_1, U_2)) \\ &= \text{Var}(g(U_1, U_2)) + \text{Var}(h(U_1, U_2)) = 1,088. \end{aligned}$$

Notons que si g est monotone croissante dans $u_1^{(1)}$ et monotone décroissante dans $u_1^{(2)}$ et h monotone croissante dans $u_2^{(1)}$ et $u_2^{(2)}$, d'après (2.5.16) le vecteur antithétique-commun $\bar{U}_2 = (\bar{U}_2^{(1)}, \bar{U}_2^{(2)}) = (1 - U_1^{(1)}, U_1^{(2)})$ est optimal.

Dans ce cas

$$E(g(U_1, U_2)h(U_1, U_2)) = E[(U_1 + e^{-U_2})5U_1(1 - U_2)^2] = 0,533$$

$$\begin{aligned} \text{Cov}(g(U_1, U_2), h(U_1, U_2)) &= E(g(U_1, U_2)h(U_1, U_2)) - E(g(U_1, U_2))E(h(U_1, U_2)) \\ &= 0,136 - (1,132)0,833 = -0,490 < 0 \end{aligned}$$

$$\text{et } \text{Var}(g(U_1, U_2) + h(U_1, U_2)) = 0,116 + 0,972 + 2(-0,490) = 0,268.$$

On remarque une réduction de la variance de 75% par rapport au cas où les variables sont indépendantes.

Considérons un cas particulier où X est de loi uniforme sur (a_1, b_1) et Y est de loi uniforme sur (a_2, b_2) ,

où $X, Y, a, b \in \mathbb{R}^n$, la fonction optimale F_{XY} peut être calculée directement.

En effet, si $g(X)$ et $h(Y)$ sont des fonctions monotones dans les directions opposées alors

$$\left(F_{X_1}(x_1), \dots, F_{X_n}(x_n) \right) = \frac{x - a_1}{b_1 - a_1} = u_1.$$

On voit que le choix de Y optimal satisfait à

$$Y = a_2 + \frac{a_1 - X}{b_1 - a_1}(b_2 - a_2).$$

Les formules ci-dessus peuvent être récapitulées dans la procédure suivante :

Procédure

- (1) Générer X selon un loi $U(a_1, b_1)$.
- (2) Calculer $A = a_2 + \frac{X - a_1}{b_1 - a_1}(b_2 - a_2)$.
- (3) Si g et h sont monotones dans la direction opposée, alors il faut prendre

$$Y = a_2 + b_2 - A.$$

- Prendre $\theta = g(X) + h(Y)$ notre estimateur désiré

Il est très important de noter que le vecteur des nombres aléatoires antithétique-commun est optimal si la dépendance est seulement permise entre les composantes de même indice des vecteurs u_1 et u_2 . Sans cette restriction il n'y a pas en général l'optimalité du vecteur des nombres aléatoires antithétique-commun.

2.5.2.3. Les fonctions g et h ne sont pas monotones

Dans cette situation il n'existe pas de procédure pratique pour résoudre le problème (2.5.2) à moins qu'on puisse trouver les expressions analytiques (ou numériques) des inverses des fonctions de distribution de $g(X)$ et $h(Y)$, qu'on peut noter par $F_{g(X)}^{-1}$ et $F_{h(Y)}^{-1}$.

Donc en appliquant la méthode de transformation inverse sur $W = g(X)$ et $Z = h(Y)$, nous obtenons

$$W = g(X) \triangleq F_{g(X)}^{-1}(U_1) = \tilde{g}(U_1),$$

et

$$Z = h(Y) \triangleq F_{h(Y)}^{-1}(U_2) = \tilde{h}(U_2).$$

Ainsi F_W^{-1} et F_Z^{-1} sont des fonctions non monotones et non décroissantes sur $(0, 1)$. On peut donc utiliser $U_2 = 1 - U_1$ pour minimiser la variance de $\tilde{g}(U_1) + \tilde{h}(U_2)$. Mais l'application de cette technique est très limitée par les difficultés d'obtenir les expressions analytiques de $\tilde{g}(U_1)$ et $\tilde{h}(U_2)$.

Maintenant nous allons décrire la procédure qui va nous permettre de réduire la variance de $g(X) + h(Y) = g^*(U_1) + h^*(U_2)$.

Considérant que cette procédure n'est pas optimale, elle n'exige pas la disponibilité de $F_{g(X)}^{-1}$ et de $F_{h(Y)}^{-1}$ mais tout en obligeant certaines informations sur le comportement de g^* et h^* , à savoir l'identification des sous régions de monotonie de g^* et h^* . Les intersections de ces régions qu'on note $L_i^{(1)} \times L_j^{(2)}$, $i = 1, \dots, m, j = 1, \dots, l$ sont alternativement disjointes, et couvrent le pavé unitaire de dimension $2n$ de sorte que g^* et h^* sont monotones dans chacune d'elles. L'idée derrière cette procédure est illustrée de la façon suivante :

Soit (U_1, U_2) un vecteur de dimension $2n$ tel que $U_1 = (U_1^{(1)}, \dots, U_1^{(n)})$ et $U_2 = (U_2^{(1)}, \dots, U_2^{(n)})$ sont deux vecteurs de dimension n avec des composantes indépendantes et chacune est distribuée uniformément sur $[0, 1]$. Le vecteur

(U_1, U_2) appartient donc à l'un des ml pavés $L_i^{(1)} \times L_j^{(2)}$, $i = 1, \dots, m, j = 1, \dots, l$ de dimension $2n$. On peut écrire

$$g^*(u_1) = \sum_{i=1}^m g^*(u_1) I_i^{(1)}(u_1), \quad h^*(u_2) = \sum_{j=1}^l h^*(u_2) I_j^{(2)}(u_2)$$

où

$$I_k^{(i)}(u) \begin{cases} 1, & \text{si } u \in L_k^{(i)}, k = 1, \dots, m \text{ ou } k = 1, \dots, l \\ 0, & \text{sinon.} \end{cases}$$

On obtient

$$\begin{aligned} E(g(X) + h(Y)) &= E(g^*(X) + h^*(Y)) \\ &= \sum_{i=1}^m \sum_{j=1}^l P_{ij} E \left[g^*(U_1) I_i^{(1)}(U_1) + h^*(U_2) I_j^{(2)}(U_2) \right], \end{aligned}$$

où

$$P_{ij} = P \left((U_1, U_2) \in L_i^{(1)} \times L_j^{(2)} \right).$$

Comme $g^*(u_1) I_i^{(1)}$ et $h^*(U_2) I_j^{(2)}$ sont des fonctions monotones dans $L_i^{(1)}$ et $L_j^{(2)}$, respectivement, la réduction de la variance est obtenue en utilisant le nombre aléatoire antithétique-commun dans chaque sous-région.

Exemple 2.5.2.

Soit g et h deux fonctions monotones par morceaux et $U_1 = (U_1^{(1)}, U_1^{(2)})$ et $U_2 = (U_2^{(1)}, U_2^{(2)})$ une paire de vecteurs avec des composantes indépendantes dont chacune est distribuée uniformément dans $[0, 1]$.

Soient

$$g(u_1^{(1)}, u_2^{(1)}) = u_1^{(1)} + u_2^{(1)},$$

$$\text{et } h(u_2^{(1)}, u_2^{(2)}) = \left((u_2^{(1)} - \frac{1}{2})^2 + (u_2^{(2)} - \frac{1}{2})^2 \right).$$

Pour des variables indépendantes nous avons

$$\text{Var}(g(U_1)) = 0,167, \quad \text{Var}(h(U_2)) = 0,011.$$

On voit aisément que les intervalles ci-dessus sont les quatre régions possibles de monotonie de h

$$\begin{aligned} L_1 &= \left[0, \frac{1}{2}\right) \times \left[0, \frac{1}{2}\right), & L_2 &= \left[0, \frac{1}{2}\right) \times \left[\frac{1}{2}, 1\right), \\ L_3 &= \left[\frac{1}{2}, 1\right) \times \left[0, \frac{1}{2}\right), & L_4 &= \left[\frac{1}{2}, 1\right) \times \left[\frac{1}{2}, 1\right). \end{aligned}$$

et on suppose g monotone sur $I = [0, 1] \times [0, 1]$, nous avons donc :

(a) $I \times L_1$, les fonctions g et h sont monotones dans la direction opposée à la fois dans u_1 et u_2 .

(b) $I \times L_2$, les fonctions g et h sont monotones dans la direction opposée dans u_1 et sont monotones dans la même la direction dans u_2 .

(c) $I \times L_3$, les fonctions g et h sont monotones dans la direction opposée dans u_2 et sont monotones dans la même la direction dans u_1 .

(d) $I \times L_4$, les fonctions g et h sont monotones dans la même la direction à la fois dans u_1 et u_2 .

D'après la procédure précédente nous avons

$$\begin{aligned} \text{(a)} \quad I \times L_1, \quad & \bar{u}_2^{(1)} = 0,5 - 0,5u_1^{(1)}, \quad \bar{u}_2^{(2)} = 0,5 - 0,5u_1^{(2)}, \\ \text{(b)} \quad I \times L_2, \quad & \bar{u}_2^{(1)} = 0,5 - 0,5u_1^{(1)}, \quad \bar{u}_2^{(2)} = 0,5 + 0,5u_1^{(2)}, \\ \text{(c)} \quad I \times L_3, \quad & \bar{u}_2^{(1)} = 0,5 + 0,5u_1^{(1)}, \quad \bar{u}_2^{(2)} = 0,5 - 0,5u_1^{(2)}, \\ \text{(d)} \quad I \times L_4, \quad & \bar{u}_2^{(1)} = 0,5 + 0,5u_1^{(1)}, \quad \bar{u}_2^{(2)} = 0,5 + 0,5u_1^{(2)}. \end{aligned}$$

Donc

$$\begin{aligned} E(g(\bar{U}_1)h(\bar{U}_2)) &= 0,186, \quad E(g(\bar{U}_1))E(h(\bar{U}_2)) = 0,167, \\ \text{et } \text{Cov}(g(\bar{U}_1), h(\bar{U}_2)) &= 0,186 - 0,167 = 0,019. \end{aligned}$$

Alors

$$\text{Var}(g(\bar{U}_1) - h(\bar{U}_2)) = 0,178 - 2(0,019) = 0,140.$$

On remarque une réduction de la variance de 21,34% relativement au cas des variables indépendantes.

2.5.3. Variables antithétiques pour des fonctions non monotones

2.5.3.1. Introduction

Supposons qu'on veut estimer

$$\theta = E \left[\sum_{i=1}^n \Psi_i(X_i) \right] \quad (2.5.17)$$

avec variance minimale, où Ψ_i sont des fonctions connues, et $F_{X_i}(x)$, $i = 1, \dots, n$, sont des fonctions de distribution connues des variables aléatoires X_i .

Généralement, la méthode des variables antithétiques consiste à induire la dépendance appropriée entre les X_i telle que la variance de l'estimateur

$$\hat{\theta} = \sum_{i=1}^n \Psi_i(X_i) \quad (2.5.18)$$

soit réduite en préservant les distributions marginales des X_i , en d'autres termes, pour une distribution marginale connue $F_{X_i}(x)$, $i = 1, \dots, n$, la méthode de variables antithétiques génère une fonction de distribution jointe $F_{X_1, \dots, X_n}(\cdot)$ pour laquelle la variance de l'équation (2.5.18) est beaucoup plus petite que la variance dans le cas où les variables X_1, \dots, X_n sont indépendantes.

Supposons que les variables X_i sont générées par la méthode de transformation inverse, alors on a :

$$X_i = F_{X_i}^{-1}(U_i) = \text{Inf} \left\{ x : F_{X_i}(x) \geq U_i \right\} \quad (2.5.19)$$

où $U_i, i = 1, \dots, n$, sont des variables aléatoires distribuées uniformément sur l'intervalle $[0, 1]$. Donc (2.5.17) s'écrit :

$$\begin{aligned} \theta &= E \left\{ \sum_i \Psi_i \left[F_{X_i}^{-1}(U_i) \right] \right\} \\ &= E \left[\sum_i g_i(U_i) \right] \quad \text{où} \quad g_i(U_i) = \Psi_i \left[F_{X_i}^{-1}(U_i) \right]. \end{aligned}$$

Ainsi, on peut résumer le problème de réduction de la variance comme suit :

Etant donné les fonctions g_1, g_2, \dots, g_n , on choisit une distribution jointe de U_1, \dots, U_n , notée F_{U_1, \dots, U_n} , pour minimiser $\text{Var}(\theta) = \text{Var}(\sum_i g_i(U_i))$. Une solution à ce problème est donnée par le théorème de Handscomb (1958) :

Théorème 2.5.2. *Si I représente l'infimum de $\text{Var}(\sum_i g_i(U_i))$ quand on considère toutes les dépendances possibles stochastiques ou fonctionnelles entre les U_i , alors pour les g_i bornées on a*

$$I = \inf_{\substack{x \in \chi \\ i=1,2,\dots,d}} \text{Var} \left(\sum_i g_i(x_i(U)) \right) \quad (2.5.20)$$

où χ est un ensemble des classes de fonctions telles que :

(1) $x(U)$ est une application biunivoque,

(2) $dx/dU = 1$ sauf en un nombre de fini points.

En d'autres termes, chaque fois que l'on a un estimateur formé d'une somme de variables aléatoires, il est possible de faire intervenir une dépendance fonctionnelle stricte telle que l'estimateur demeure sans biais, tandis que sa variance devient proche de la plus faible qui puisse être obtenue à l'aide de ces variables.

On note ici que Wilson (1979) a travaillé sur ce théorème avec des fonctions g_1, \dots, g_n non bornées et avec variances finies. Malheureusement, les deux théorèmes de Wilson (1979) et de Handscomb (1958) ne traitent pas l'existence de la valeur minimale exacte de la Variance de $(\hat{\theta})$, et on ne sait pas si l'infimum peut être atteint et si cela peut se faire en utilisant une seule variable aléatoire U malgré que U apparaît dans chaque $x_i(U)$ de (2.5.20).

C'est dans le souci de répondre positivement à ces deux problèmes qu'on parle maintenant de la version modifiée du théorème de Handscomb.

2.5.3.2. La version modifiée du théorème de Handscomb

Théorème 2.5.3. *Soient $g_1(U_1), g_2(U_2), \dots, g_n(U_n)$ des fonctions mesurables et bornées sur l'intervalle $[0, 1]$ et Ξ_n l'ensemble de toutes les distributions jointes possibles sur $[0, 1]^n$ avec des marginales uniformes. Alors il existe des transformations $h_1(U_1), h_2(U_2), \dots, h_n(U_n)$ telles que :*

$$\begin{aligned} & \min_{F \in \Xi_n} \text{Var} \left[g_1(U_1) + g_2(U_2) + \dots + g_n(U_n) \right] \\ &= \text{Var} \left[g_1(h_1(U)) + g_2(h_2(U)) + \dots + g_n(h_n(U)) \right] \end{aligned}$$

où $(h_i(U))$ sont de lois uniformes sur $(0, 1)$ pour $i = 1 \dots, n$.

On note ici que c'est le même U qui apparaît dans tous les $h_k(\cdot)$, $k = 1, 2, \dots, n$ et on note F^* la fonction de distribution optimale.

On trouve la démonstration de ce théorème dans Rubinstein & Samoradnitsky (1986) et le théorème suivant est la généralisation du théorème (2.5.3).

Théorème 2.5.4. *Soit $\varphi[g_1(U_1), g_2(U_2), \dots, g_n(U_n)] : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continue définie sur un ensemble borné et fermé, où $U_i = (U_i^1, \dots, U_i^m)$, $i = 1, 2, \dots, n$, est un vecteur de dimension m avec des composantes indépendantes chacune distribuée selon une loi uniforme sur $(0, 1)$ et $g_1(\cdot), \dots, g_n(\cdot)$ des fonctions mesurables bornées sur $[0, 1]^m$.*

Soit Ξ_n^m un ensemble des toutes les distributions jointes sur $[0, 1]^{n \times m}$ avec des marginales uniformes de dimension m . Alors il existe des transformations $h_i^j(U)$, $i = 1, \dots, n, j = 1, \dots, m$, telle que :

$$\begin{aligned}
& \min_{F \in \Xi_n^m} E\varphi \left[g_1(U_1^1, \dots, U_1^m), \dots, g_n(U_n^1, \dots, U_n^m) \right] \\
&= E\varphi \left[g_1(h_1^1(U_1), \dots, h_1^m(U_m)), \dots, g_n(h_n^1(U_1), \dots, h_n^m(U_m)) \right],
\end{aligned}$$

où

$$h_i^j(U^j) \sim U(0, 1), \quad i = 1, \dots, n, j = 1, \dots, m, \quad (2.5.21)$$

$$\begin{aligned}
& h_i^j(U^j) \text{ ne dépend pas de } h_i^k(U^k), \\
& k \neq j, k = 1, \dots, m, i = 1, \dots, n.
\end{aligned} \quad (2.5.22)$$

Corollaire 2.5.1. *Supposons qu'on veut estimer*

$$\theta = E \left[\sum_{i=1}^n g_i(U_i) \right] \quad (2.5.23)$$

avec variance minimale.

Notons

$$\varphi \left[g_1(U_1), g_2(U_2), \dots, g_n(U_n) \right] = \sum_{i=1}^n \left[g_i(U_i) - \theta \right]^2. \quad (2.5.24)$$

Alors par le théorème précédent il existe des transformations $h_i(U)$ qui satisfont (2.5.21) et (2.5.22) telles que :

$$\min_{F \in \Xi_n} \text{Var}[g_1(U_1) + g_2(U_2) + \dots + g_n(U_n)] \quad (2.5.25)$$

$$= \text{Var}[g_1(h_1(U)) + g_2(h_2(U)) + \dots + g_n(h_n(U))],$$

$$\text{où } (h_i(U) = (h_i^1(U^1), \dots, h_i^m(U^m)), i = 1, \dots, n.$$

2.5.3.3. Procédure optimale pour $n = 2$

On remarque que les deux théorèmes précédents sont des résultats d'existence, nulle part n'est mentionné la façon de construire la transformation optimale $h_1(U), h_2(U), \dots, h_n(U)$ qui minimise $\text{Var}(\sum_{i=1}^n [g_i(U_i)])$.

Dans cette section nous allons présenter deux procédures pratiques pour construire $h_1(U), h_2(U)$ (pour $n = 2$) c'est-à-dire, deux procédures pour générer une distribution jointe F_{U_1, U_2} qui minimise $\text{Var}[g_1(U_1) + g_2(U_2)]$ où g_1 et g_2 sont deux fonctions non-monotones.

(a) Résultats théoriques

Soit $g(u)$ une fonction quelconque mesurable sur $[0, 1]$. Soit $F_g(x) = P[g(U) \leq x]$ une fonction de distribution de la variable aléatoire $g(U)$. Par la méthode de la transformation inverse $F_g^{-1}(U)$ est distribuée comme $g(U)$. Il est clair que $F_g^{-1}(1 - U)$ est aussi distribuée comme $g(U)$. Ici $F_g^{-1}(U)$ et $F_g^{-1}(1 - U)$ sont des fonctions non décroissantes et non croissantes de U respectivement.

On note $g^*(u) = F_g^{-1}(u)$ et $g_*(u) = F_g^{-1}(1 - u)$ et on les appelle donc des fonctions non décroissantes et non croissantes de U distribuées selon $g(U)$.

Proposition 2.5.2. *Pour chaque fonction mesurable g sur $[0, 1]$ il existe une paire de transformations $\{\bar{h}_g(u), \underline{h}_g(u)\}$ telles que*

- (1) $\bar{h}_g(U)$ est de loi uniforme sur $(0, 1)$ et $\underline{h}_g(U)$ est de loi uniforme sur $(0, 1)$ et
- (2) $g^*[\bar{h}_g(U)] = g(U)$ presque sûrement sur $[0, 1]$ et $g_*[\underline{h}_g(U)] = g(U)$ presque sûrement sur $[0, 1]$.

Exemple 2.5.3. *Soit*

$$g(u) = \begin{cases} 4u, & \text{si } 0 < u \leq \frac{1}{4}, \\ 0, 5, & \text{si } \frac{1}{4} < u \leq \frac{1}{2}, \\ 4u - 2, & \text{si } \frac{1}{2} < u \leq \frac{3}{4}, \\ 0, 5, & \text{si } \frac{3}{4} < u \leq 1. \end{cases}$$

Alors

$$F_g(x) = \begin{cases} \frac{1}{2}x, & \text{si } 0 < x < \frac{1}{2}, \\ \frac{1}{2}x + 0, 5. & \text{si } \frac{1}{2} \leq x \leq 1. \end{cases}$$

Ainsi

$$g^*(u) = F_g^{-1}(u) = \begin{cases} 2u, & \text{si } 0 < u < \frac{1}{4}, \\ 0,5, & \text{si } \frac{1}{4} \leq u \leq \frac{3}{4}, \\ 2u - 1. & \text{si } \frac{3}{4} < u \leq 1. \end{cases}$$

Par conséquent on a

$$g_*(u) = F_g^{-1}(1 - u) = \begin{cases} 1 - 2u, & \text{si } 0 < u < \frac{1}{4}, \\ 0,5, & \text{si } \frac{1}{4} \leq u \leq \frac{3}{4}, \\ 2 - 2u, & \text{si } \frac{3}{4} < u \leq 1. \end{cases}$$

Par la proposition (2.5.2), il existe une paire de transformations $\{\bar{h}_g, \underline{h}_g\}$ qui sont fonction d'une simple variable U .

Par exemple l'une des fonctions possibles $\underline{h}_g(u)$ est :

$$\underline{h}_g(u) = \begin{cases} 1 - 2u, & \text{si } 0 < u < \frac{1}{8}, \\ 0,5 - u, & \text{si } \frac{1}{8} \leq u < \frac{1}{4}, \\ 1 - u, & \text{si } \frac{1}{4} \leq u < \frac{1}{2}, \\ 2 - 2u, & \text{si } \frac{1}{2} \leq u < \frac{5}{8}, \\ 1,5 - 2u, & \text{si } \frac{5}{8} \leq u < \frac{3}{4}, \\ \frac{5}{4} - u, & \text{si } \frac{3}{4} \leq u \leq 1. \end{cases}$$

(b) Procédure optimale et exemple

Maintenant nous allons utiliser la proposition (2.5.2) pour construire une paire de transformations $\{h_1(u), h_2(u)\}$ qui génère la fonction de distribution optimale et par conséquent minimise

$$\text{Var}[g_1(U_1) + g_2(U_2)]. \quad (2.5.26)$$

On commence avec le cas où l'une des fonctions, disons g_2 , est monotone.

Lemme 2.5.1. *Si $g_1(u_1)$ et $g_2(u_2)$ sont deux fonctions mesurables sur $[0, 1]$ et $g_2(u_2)$ est monotone, alors il existe une transformation $h(u)$ telle que*

$$(1) \ h(U) \sim U_{(0,1)} \text{ et}$$

$$(2) \ \min_{F \in \Xi_2} \{ \text{Var}_F[g_1(U_1) + g_2(U_2)] \} = \text{Var}[g_1(U) + g_2(h(U))].$$

Démonstration

Supposons que $g_2(u)$ est non croissante dans u . Par la proposition il existe une mesure qui préserve la transformation $\bar{h}_{g_1}(u_1)$ telle que

$$g_1^*[\bar{h}_{g_1}(u_1)] = g_1(u_1) \quad \text{presque sûrement.}$$

Pour démontrer (1), on prend $h(u) = \bar{h}_{g_1}(u)$.

Pour démontrer (2), on suppose que $F(u_1, u_2)$ est une fonction de distribution quelconque appartenant à l'ensemble Ξ_2 de toutes les distributions possibles.

$$\min_{F \in \Xi_2} \text{Var}_F[g_1(U_1) + g_2(U_2)] = \min_{F \in \Xi_2} \text{Var}_F[g_1^*(\bar{h}_{g_1}(U_1)) + g_2(U_2)]$$

$$\geq \text{Var}[g_1^*(U) + g_2(U)] = \text{Var}[g_1^*(\bar{h}_{g_1}(U_1)) + g_2(\bar{h}_{g_1}(U))]$$

$$= \text{Var}[g_1(U) + g_2(\bar{h}_{g_1}(U))] = \text{Var}[g_1(U) + g_2(U)]. \quad (2.5.27)$$

L'inégalité provient de $g_1^*(u) = F_{g_1}^{-1}(u)$ et $g_2(u_2)$ sont monotones dans les directions opposées. La première égalité après l'inégalité provient du fait que $\bar{h}_{g_1}(U)$ est de loi uniforme sur $(0, 1)$ et les deux dernières égalités dans l'équation (2.5.27) sont évidentes.

Exemple 2.5.4.

Considérons le cas où g_2 est une fonction monotone croissante sur $[0, 1]$ et g_1 est une fonction symétrique et unimodale telle que $g_1(0, 5 - \theta) = g_1(0, 5 + \theta)$, pour tout $\theta \in [0, 1/2]$ et g_1 est monotone sur $[0, 1/2]$, par exemple, non croissante.

Alors on trouve

$$\bar{h}_{g_1}(u) = \begin{cases} 1 - 2u, & \text{si } 0 < u \leq 0.5, \\ 2u - 1, & \text{si } 0.5 \leq u \leq 1. \end{cases}$$

D'après la propriété (2) du lemme 2.5.1 nous avons

$$\min_{F_{u_1 u_2} \in \Xi_2} \{\text{Var}[g_1(U_1) + g_2(U_2)]\} = \text{Var}[g_1(U) + g_2(\bar{h}_{g_1}(U))].$$

Comme exemple numérique, soit $g_1(u_1) = \sin(\pi u_1)$ et $g_2(u_2) = u_2^2$.

Posons

$$\hat{\theta}_1 = g_1(U_1) + g_2(U_2), \quad \hat{\theta}_2 = g_1(U_1) + g_2(\bar{h}_{g_1}(U_2)).$$

Alors, par la méthode de Monte Carlo et par la procédure optimale, nous obtenons respectivement

$$\text{Var}(\hat{\theta}_1) = 0,1836, \quad \text{Var}(\hat{\theta}_2) = 0,0189.$$

Le gain dans la réduction de la variance est

$$\epsilon = \frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)} = 9,71.$$

Si g_1 est une fonction continue par morceau, g_2 une fonction monotone et si la fonction de distribution de la variable aléatoire $g_1(U)$ est continue alors la procédure suivante est efficace pour minimiser la variance de (2.5.26).

Procédure

- (1) Générer U_1 selon une loi uniforme sur $(0, 1)$.
- (2) Trouver l'ensemble $\{u \in [0, 1] : g_1(u) \leq g_1(u_1)\}$ qui est formé d'une union de quelques intervalles.
- (3) Calculer l qui est la somme des longueurs de ces intervalles.

(4) D  duire $\theta_2 = g_1(u_1) + g_2(l)$.

Pour v  rifier si cette proc  dure minimise (2.5.26), on d  finit $g_2(l)$ comme suit :

$$g_2(l) = g_2\{P[g_1(U)] \leq g_1(u_1)\} = g_2[F_{g_1}(g_1(u_1))] = g_2[\bar{h}_{g_1}(u_1)].$$

Consid  rons maintenant le cas o   g_1 et g_2 sont des fonctions mesurables quelconques. On va montrer comment la proposition est utile pour construire une paire de transformations $\{h_1(U), h_2(U)\}$ (pour g  n  rer une fonction de distribution optimale) qui minimise (2.5.26).

Consid  rons premi  rement une paire de vecteurs $\{u, \bar{h}_{g_1}(u)\}$ et $\{u, \underline{h}_{g_2}(u)\}$ qui d  finissent une paire de distributions sur $[0, 1] \times [0, 1]$, avec des marginales uniformes qu'on note $G_1(U^{(1)}, U^{(2)})$ et $G_2(U^{(1)}, U^{(2)})$. Alors la proc  dure suivante minimise (2.5.26) :

Proc  dure

- (1) G  n  rer U selon une loi uniforme sur $(0, 1)$.
- (2) Trouver les distributions conditionnelles
 $G_1(U^{(1)}|U^{(2)} = U)$ et $G_2(U^{(1)}|U^{(2)} = U)$.
- (3) G  n  rer \tilde{U}_1 et \tilde{U}_2 , respectivement,    partir de
 $G_1(\cdot|U^{(1)} = U)$ et $G_2(\cdot|U^{(2)} = U)$.
- (4) Calculer $g_1(\tilde{U}_1) + g_2(\tilde{U}_2)$ notre estimateur d  sir  .

Pour voir si cette proc  dure minimise (2.5.26) (optimal), on note premi  rement que la distribution $F_{\tilde{U}_1, \tilde{U}_2}$ poss  dent des marginales uniformes puisque les distributions G_1 et G_2 en admettent. De plus pour la construction de la proc  dure, il faut tenir compte que

$$\bar{h}_{g_1}(\tilde{U}_1) = U \text{ et } \underline{h}_{g_2}(\tilde{U}_2) = U. \text{ Ainsi } g_1(\tilde{U}_1) = g_1^*(U) \text{ et } g_2(\tilde{U}_2) = g_*(U).$$

$$\begin{aligned}
& \text{Donc} \quad \min_{F_{U_1 U_2} \in \Xi_2} \text{Var}[g_1(U_1) + g_2(U_2)] \\
&= \min_{F_{U_1 U_2} \in \Xi_2} \text{Var}[g_1^*(\bar{h}_{g_1}(U_1)) + g_{2*}(\underline{h}_{g_2}(U_2))], \\
&\geq \text{Var}[g_1^*(U) + g_{2*}(U)], \\
&= \text{Var} \left\{ g_1^*[\bar{h}_{g_1}(\tilde{U}_1)] + g_{2*}[\underline{h}_{g_2}(\tilde{U}_2)] \right\}, \\
&= \text{Var}[g_1(\tilde{U}_1) + g_2(\tilde{U}_2)].
\end{aligned}$$

Donc dans tout l'ensemble des fonctions de distributions bi-dimensionnelles, la fonction de distribution $F_{\tilde{U}_1, \tilde{U}_2}$ minimise (2.5.26).

Remarque

Si les fonctions de distribution de g_1 et g_2 sont continues, il est facile de voir que la procédure provient de la correspondance bi-univoque entre les valeurs de g_1 et g_2 .

Chapitre 3

NOUVELLE MÉTHODE DE SIMULATION AVEC VARIABLES ANTITHÉTIQUES

Ce chapitre va contenir d'autres nouvelles méthodes de simulation avec variables antithétiques sur des fonctions monotones, non monotones et en plus nous allons introduire la fonction instrumentale sur les densités qui sont difficilement simulables.

3.1. FONCTIONS MONOTONES

3.1.1. Principe général et exemple

L'objectif de cette méthode est d'évaluer l'intégrale

$$I = \int_0^1 f(x)dx$$

où f est une fonction croissante sur l'intervalle $(0, 1]$ et la méthode que nous allons proposer construit une série de subdivisions de l'intervalle $(0, 1]$ en plusieurs sous-intervalles de même longueur. Les différentes étapes de subdivision sont décrites de la façon suivante :

3.1.1.1. Cas où $m = 1$

- (1) **Etape 0 :** On considère $n = ml$ où m est le nombre de sections dans l'intervalle $(0, 1]$ dont chacune est subdivisée en l sous-intervalles. Supposons $m = 1$ donc $n = l$ ce qui revient à subdiviser l'intervalle $(0, 1]$ en n

sous-intervalles I_1, I_2, \dots, I_n de même longueur et chaque I_i , $i = 1, 2, \dots, n$ est de la forme

$$I_i = \left(\frac{i-1}{n}, \frac{i}{n} \right]. \quad (3.1.1)$$

(2) **Etape 1** : Chaque I_{i_1} ($i_1 = 1, 2, \dots, n$) est subdivisé en n sous-intervalles I_{i_1, i_2} de même longueur avec $i_1, i_2 = 1, \dots, n$ et ils seront de la forme

$$I_{i_1 i_2} = \left(\frac{i_1-1}{n} + \frac{i_2-1}{n^2}, \frac{i_1-1}{n} + \frac{i_2}{n^2} \right]. \quad (3.1.2)$$

(3) **Etape k** : À la k -ième étape nous aurons des sous-intervalles $I_{i_1 i_2 \dots i_{k+1}}$ avec $i_1, i_2, \dots, i_{k+1} = 1, \dots, n$ et ils seront de la forme suivante :

$$I_{i_1 i_2 \dots i_{k+1}} = \left(\frac{i_1-1}{n} + \frac{i_2-1}{n^2} + \dots + \frac{i_{k+1}-1}{n^{k+1}}, \frac{i_1-1}{n} + \dots + \frac{i_k-1}{n^k} + \frac{i_{k+1}}{n^{k+1}} \right].$$

On choisit n sous intervalles dans chaque section parmi $\{I_{i_1 i_2 \dots i_{k+1}} : i_1, i_2, \dots, i_{k+1} = 1, \dots, n\}$ à l'aide de k permutations, c'est-à-dire qu'on choisit les sous-intervalles $I_{\pi^{(k)} i}$

$$I_{1\pi_1^1\pi_1^2\dots\pi_1^k}, I_{2\pi_2^1\pi_2^2\dots\pi_2^k}, \dots, I_{n\pi_n^1\pi_n^2\dots\pi_n^k}$$

où on définit $\pi = \left\{ \pi^j \right\}_{j=1}^{\infty}$ une suite de permutations *iid* uniforme sur $\{1, 2, \dots, n\}$

avec $\pi^{(k)} = \left\{ \pi_i^j \right\}_{j=1}^k$, $k \geq 1$ et $i = 1, \dots, n$ les k premières permutations et $\pi^0 = \{1, 2, \dots, n\}$ est la permutation identité.

Dans chaque intervalle $I_{i\pi^{(k)}}$ on génère uniformément les points $U_{i\pi^{(k)}}$ et on définira l'estimateur $\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^n f(U_{i\pi^{(k)}})$ où f est une fonction croissante.

La loi $\mathcal{L}(U_{i\pi^{(k)}} | \pi^{(k)})$ est une loi uniforme sur $I_{i\pi^{(k)}}$ et $\mathcal{L}(U_{i\pi^{(k)}})$ est de loi marginale $\mathcal{U}\left(\left(\frac{i-1}{n}, \frac{i}{n}\right]\right)$.

3.1.1.2. Généralisation : cas où $m > 1$

On sépare $(0, 1]$ en m sous-intervalles de même longueur appelés sections. On pose

$$\mu = \frac{1}{m} \sum_{j=1}^m \mu_j, \quad (3.1.3)$$

avec

$$\mu_j = \int_0^1 f_j(x) dx \quad (3.1.4)$$

$$\text{où } f_j(x) = f\left((x + j - 1)/m\right).$$

On considère $n = ml$ où l est le nombre de sous-intervalles dans chaque section. Chacun des I_j , $j = 1, \dots, m$ est estimé selon l'approche décrite précédemment basée sur l observations. En pratique on aura

$$\hat{\mu} = \frac{1}{n} \sum_{j=1}^m \sum_{k=1}^l f(x_{(j-1)l+k})$$

avec

$$x_{(j-1)l+1}, \dots, x_{(j-1)l+l} \in \left(\frac{j-1}{m}, \frac{j}{m}\right], \quad j = 1, \dots, m.$$

Exemple 3.1.1.

Donnons-nous un exemple pour $n = 10$, $m = 2$, $l = 5$, $k = 1$:

Etape 0 ($k = 0$) :

Génération des uniformes sur $(0, 1]$

1-ère section : 0,8298889 0,9109330 0,2339573 0,9713439 0.8197696

2-ième section : 0,5705076 0,9224039 0,6984488 0,4595329 0,5064768.

Les bornes inférieures des quatres sous-intervalles I_i de chaque section sont :

1ère section : 0,0 0,1 0,2 0,3 0,4

2ième section : 0,5 0,6 0,7 0,8 0,9.

Les points générés dans chaque sous-intervalle I_i sont :

1-ère section : 0,08298889 0,1910933 0,2233957 0,3971344 0,4819770

2-ième section : 0,55705076 0,6922404 0,7698449 0,8459533 0,9506477.

Etape 1 :

- On subdivise chaque I_i de chaque section en 5 sous-intervalles trouvés à l'étape 0.

- Ensuite on génère cinq permutations dans chaque section ce qui nous permet de choisir cinq sous intervalles parmi $5^5 = 25$.

Permutations :

1ère section : 3 5 1 4 2,

2ième section : 2 3 4 5 1.

Dans l'intervalle $(0, 1]$ on choisit dix sous-intervalles parmi les cinquante sous-intervalles à l'aide de dix permutations dont cinq dans chaque section :

1-ère section : 0,04 0,18 0,20 0,36 0,42,

2-ième section : 0,52 0,64 0,76 0,88 0,90.

Les points obtenus qui sont dans chaque sous-intervalle précédent sont :

1-ère section : 0,04829889 0,1891093 0,2023396 0,3697134 0,4281977,

2-ième section : 0,52570508 0,6492240 0,7669845 0,8845953 0,9050648.

Ainsi l'estimateur cherché est égal à

$$\mu_1 = \frac{1}{10} \left\{ f(0.04829889) + f(0.1891093) + \cdots + f(0.9050648) \right\}.$$

3.1.2. Formules des variances pour la fonction identité et exponentielle

Dans cette partie on donne deux tableaux dont le premier contient les formules des variances dans le cas du tirage stratifié à l'étape 0 et à l'étape 1 et l'autre contient les formules du premier tableau appliquées sur deux cas particuliers : cas uniforme et cas exponentiel. Les développements de ces résultats sont donnés dans la sous-section (3.1.4).

Dans le cas du tirage simple la variance est égale

$$\frac{\sigma^2}{n} = \frac{\text{Var}(f(U))}{n}.$$

Tirage	Etape 0	Etape 1
Tirage stratifié	$\text{Var}(\hat{\mu}_0) = \frac{\sigma^2}{n} - \frac{1}{n^2} \sum_i (\mu_i - \mu)^2$ <p>où $\mu_i = E\left(f(U) U \in I_i\right)$ et $\mu = E\left(f(U) U \in (0, 1]\right)$</p>	$\text{Var}(\hat{\mu}_1) = \text{Var}(\hat{\mu}_0).$ $- \frac{1}{n^2} \frac{1}{n(n-1)} \sum_l \sum_{i \neq j} (\mu_{il} - \mu_i)(\mu_{jl} - \mu_j)$ <p>où $\mu_{il} = E\left(f(U) U \in I_{il}\right).$</p>

TABLE 3.1. Tableau 1 des formules des variances : cas stratifié,
 $m = 1$

Pour la fonction identité $f(x) = x$, la variance dans le cas du tirage simple est égale à :

$$\frac{\sigma^2}{n} = \frac{\text{Var}(f(U))}{n} = \frac{\text{Var}(U)}{n} = \frac{1}{12n},$$

tandis que pour la fonction exponentielle $f(x) = \exp(tx)$, la variance dans le cas du tirage simple est égale à :

$$\begin{aligned} \frac{\sigma^2}{n} &= \frac{\text{Var}(f(U))}{n} = \frac{\text{Var}(e^{tU})}{n} \\ &= \frac{1}{2nt^2} \left(e^t - 1 \right) \left[\left(t + 2 \right) + \left(t - 2 \right) e^t \right]. \end{aligned}$$

Le tableau suivant contient les formules de la variance à l'étape 0 et à l'étape 1 pour les fonctions identité et exponentielle pour le cas stratifié.

Fonction	Etape 0	Etape 1
f(x)=x	$\text{Var}(\hat{\mu}_0) = \frac{\sigma^2}{n} - \frac{1}{n^2} \sum_i (\mu_i - \mu)^2.$ $\text{Par calcul } \text{Var}(\hat{\mu}_0) = \left[\frac{1}{12n} \frac{1}{n^2} \right].$ $\text{avec } \frac{1}{n^2} \sum_i (\mu_i - \mu)^2 = \frac{1}{n^2} \left[\frac{(n+1)(n-1)}{12n} \right]$ $\text{et } \mu_i = \frac{2i-1}{n}.$ $\mu = \frac{1}{2}.$	$\text{Var}(\hat{\mu}_1) = \text{Var}(\hat{\mu}_0) - \sum_i \sum_j \frac{1}{n^2} (\mu_{ij} - \mu_i)^2.$ $\text{Par calcul } \text{Var}(\hat{\mu}_1) = \left[\frac{1}{12n} \frac{1}{n^4} \right]$ $\sum_i \sum_j \frac{1}{n^2} (\mu_{ij} - \mu_i)^2 = \frac{1}{n^2} \left[\frac{(n+1)(n-1)}{12n} \right]$ $\text{et } \mu_{ij} = \frac{2(i-1)}{n} - \frac{2j-1}{n^2}$
f(x)=e ^{tx}	$\text{Var}(\hat{\mu}_0) = \frac{\sigma^2}{n} - \frac{1}{n^2} \sum_i (\mu_i - \mu)^2.$ $\text{Par calcul } \text{Var}(\hat{\mu}_0) = \left[\frac{t(e^t-1)(1+e^t)}{24n^3} \right]$ $\text{et } \frac{1}{n^2} \sum_i (\mu_i - \mu)^2$ $= \frac{1}{nt^2} (e^t - 1) \left[\frac{n(e^t+1)(e^{\frac{t}{n}}-1)}{(e^{\frac{t}{n}}+1)} - (e^t - 1) \right]$ $\text{où } \mu_i = n \int_{\frac{i-1}{n}}^{\frac{i}{n}} e^{tu} du$ $= \frac{n}{t} \left[e^{\frac{t(i-1)}{n}} \left(e^{\frac{t}{n}} - 1 \right) \right]$ $\mu = \frac{1}{t} (e^t - 1).$	$\text{Var}(\hat{\mu}_1) = \text{Var}(\hat{\mu}_0)$ $- \frac{1}{n^2} \frac{1}{n(n-1)} \sum_l \sum_{i \neq j} (\mu_{il} - \mu_i)(\mu_{jl} - \mu_j).$ $\text{Or } \frac{1}{n^2} \frac{1}{n(n-1)} \sum_l \sum_{i \neq j} (\mu_{il} - \mu_i)(\mu_{jl} - \mu_j)$ $= \frac{1}{24n^3} (e^t - 1) \left[\frac{2n(e^t-1)-t(e^t+1)}{n-1} \right].$ $\text{Var}(\hat{\mu}_1) = \frac{1}{24n^3} (e^t - 1) \left[(t-2)e^t + (2+t) \right]$ $\mu_{il} = n^2 \int_{\frac{i-1}{n} + \frac{l-1}{n^2}}^{\frac{i-1}{n} + \frac{l}{n^2}} e^{tu} du$ $= \frac{n^2}{t} e^{t(\frac{i-1}{n} + \frac{l-1}{n^2})} \left(e^{\frac{t}{n^2}} - 1 \right).$

TABLE 3.2. Tableau 2 : Résumé des formules des variances : Cas Stratifié pour $f(x) = x$ et $f(x) = e^{tx}$ ($m = 1$)

3.1.3. Résultats théoriques

Théorème 3.1.1. *Si on pose $\mu_{0,i} = E(f(U_{\pi(0)i}))$ avec $U_{\pi(0)i} \sim \mathcal{U}\left(\frac{i-1}{n}, \frac{i}{n}\right)$ et $\sigma_{0,i}^2 = \text{Var}[f(U_{\pi(0)i})]$ alors l'estimateur $\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^n f(U_{\pi(k)i})$ est sans biais et $\text{Var}(\hat{\mu}_k) \leq \text{Var}(\hat{\mu}_{k-1})$ pour $k \geq 1$.*

Démonstration

Ici, $\hat{\mu}_k$ est un estimateur sans biais car

$$\begin{aligned} E(\hat{\mu}_k) &= \frac{1}{n} \sum_{i=1}^n E(f(U_{\pi(k)i})) \\ &= \frac{1}{n} \sum_{i=1}^n \mu_i = \mu, \end{aligned}$$

$$\text{où } \mu_i = E(f(U_{\pi(k)i})).$$

Démontrons que $\text{Var}(\hat{\mu}_k) \leq \text{Var}(\hat{\mu}_{k-1})$. On a

$$\begin{aligned} \text{Var}(\hat{\mu}_k) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n f(U_{\pi(k)i})\right), \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(f(U_{\pi(k)i})) + \frac{2}{n^2} \sum_{i_1 \leq i_2} \text{Cov}(f(U_{\pi(k)i_1}), f(U_{\pi(k)i_2})), \\ &= \frac{1}{n^2} \left\{ \sum_{i=1}^n \sigma_i^2 + 2 \sum_{i_1 \leq i_2} \text{Cov}(f(U_{\pi(k)i_1}), f(U_{\pi(k)i_2})) \right\}. \end{aligned}$$

Calculons

$$\begin{aligned} \text{Cov}(f(U_{\pi(k)i_1}), f(U_{\pi(k)i_2})) &= E\left(\text{Cov}(f(U_{\pi(k)i_1}), f(U_{\pi(k)i_2})) | \pi^{(k)}\right) \\ &+ \text{Cov}\left(E(f(U_{\pi(k)i_1}) | \pi^{(k)}), E(f(U_{\pi(k)i_2}) | \pi^{(k)})\right), \end{aligned} \quad (3.1.5)$$

où

$$\text{Cov}(f(U_{\pi^{(k)}i_1}), f(U_{\pi^{(k)}i_2})|\pi^{(k)}) = 0,$$

car conditionnellement au choix de $\pi^{(k)}$, $U_{\pi^{(k)}i_1}$ et $U_{\pi^{(k)}i_2}$ sont des variables uniformes et indépendantes.

Il reste dans l'expression (3.1.5)

$$\text{Cov}(f(U_{\pi^{(k)}i_1}), f(U_{\pi^{(k)}i_2})) = \text{Cov}\left(E(f(U_{\pi^{(k)}i_1}|\pi^{(k)})), E(f(U_{\pi^{(k)}i_2}|\pi^{(k)}))\right),$$

donc

$$\begin{aligned} & \text{Cov}\left(E(f(U_{\pi^{(k)}i_1}|\pi^{(k)})), E(f(U_{\pi^{(k)}i_2}|\pi^{(k)}))\right) \\ &= E\left[\text{Cov}\left\{E\left(f(U_{\pi^{(k)}i_1}|\pi^{(k)})\right), E\left(f(U_{\pi^{(k)}i_2}|\pi^{(k)})\right)|\pi^{(k-1)}\right\}\right], \\ &+ \text{Cov}\left[E\left\{E\left(f(U_{\pi^{(k)}i_1}|\pi^{(k)})\right)|\pi^{(k-1)}\right\}, E\left\{E\left(f(U_{\pi^{(k)}i_2}|\pi^{(k)})\right)|\pi^{(k-1)}\right\}\right]. \end{aligned}$$

Maintenant,

$$E\left\{E\left(f(U_{\pi^{(k)}i_1}|\pi^{(k)})\right)|\pi^{(k-1)}\right\} = E\left\{f(U_{\pi^{(k)}i_1}|\pi^{(k-1)})\right\},$$

et

$$E\left\{E\left(f(U_{\pi^{(k)}i_2}|\pi^{(k)})\right)|\pi^{(k-1)}\right\} = E\left\{f(U_{\pi^{(k-1)}i_2}|\pi^{(k-1)})\right\}.$$

Notons

$$\mu_i(\pi^{(k)}) = E \left[f(U_{\pi^{(k)} i_1}) | \pi^{(k)} \right],$$

$$\mu_{i,j}(\pi^{(k)}) = E \left[f(U_{\pi^{(k)} i_1}) | \pi^{(k+1)}, \pi_i^{(k+1)} = j \right].$$

Ainsi,

$$\begin{aligned} & \text{Cov} \left\{ E \left(f(U_{\pi^{(k)} i_1}) | \pi^{(k)} \right), E \left(f(U_{\pi^{(k)} i_2}) | \pi^{(k)} \right) \right\} \\ &= \frac{1}{n(n-1)} \sum_{j_1 \neq j_2} \sum \left[\left\{ \left(\mu_{i_1, j_1}(\pi^{(k-1)}) - \mu_{i_1}(\pi^{(k-1)}) \right) \right\} \left\{ \left(\mu_{i_2, j_2}(\pi^{(k-1)}) - \mu_{i_2}(\pi^{(k-1)}) \right) \right\} \right], \\ &= \frac{1}{n(n-1)} \sum_{j_1} \sum_{j_2} \left[\left\{ \left(\mu_{i_1, j_1}(\pi^{(k-1)}) - \mu_{i_1}(\pi^{(k-1)}) \right) \right\} \left\{ \left(\mu_{i_2, j_2}(\pi^{(k-1)}) - \mu_{i_2}(\pi^{(k-1)}) \right) \right\} \right] \\ &\quad - \frac{1}{n(n-1)} \sum_j \left[\left\{ \left(\mu_{i_1, j}(\pi^{(k-1)}) - \mu_{i_1}(\pi^{(k-1)}) \right) \right\} \left\{ \left(\mu_{i_2, j}(\pi^{(k-1)}) - \mu_{i_2}(\pi^{(k-1)}) \right) \right\} \right], \\ &= \frac{1}{n(n-1)} \left\{ \sum_j \left(\mu_{i_1, j}(\pi^{(k-1)}) - \mu_{i_1}(\pi^{(k-1)}) \right) \right\} \left\{ \sum_j \left(\mu_{i_2, j}(\pi^{(k-1)}) - \mu_{i_2}(\pi^{(k-1)}) \right) \right\} \\ &\quad - \frac{1}{n(n-1)} \sum_j \left\{ \left(\mu_{i_1, j}(\pi^{(k-1)}) - \mu_{i_1}(\pi^{(k-1)}) \right) \left(\mu_{i_2, j}(\pi^{(k-1)}) - \mu_{i_2}(\pi^{(k-1)}) \right) \right\}. \end{aligned}$$

Notons que,

$$\left(\sum \mu_{l,j}(\pi^{(k-1)}) - \mu_l(\pi^{(k-1)}) \right) = 0 \quad \text{pour tout } l.$$

Posons

$$h_1(j) = \mu_{i_1,j} \quad \text{et} \quad h_2(j) = \mu_{i_2,j}.$$

Considérons J de loi uniforme discrète sur $(\{1, 2, \dots, n\})$. Ainsi, h_1 et h_2 croissent avec J d'où

$$\frac{1}{n-1} \text{Cov}\left(h_1(J), h_2(J)\right) \geq 0.$$

Ainsi

$$\begin{aligned} \text{Cov}(f(U_{\pi^{(k)}i_1}), f(U_{\pi^{(k)}i_2})) &= \text{Cov}(f(U_{\pi^{(k-1)}i_1}), f(U_{\pi^{(k-1)}i_2})) \\ &- \frac{1}{n(n-1)} \sum_j \left\{ \left(\mu_{i_1,j}(\pi^{(k-1)}) - \mu_{i_1} \right) \left(\mu_{i_2,j}(\pi^{(k-1)}) - \mu_{i_2} \right) \right\}. \end{aligned}$$

Or

$$\frac{1}{n(n-1)} \sum_j \left\{ \left(\mu_{i_1,j}(\pi^{(k-1)}) - \mu_{i_1} \right) \left(\mu_{i_2,j}(\pi^{(k-1)}) - \mu_{i_2} \right) \right\} \geq 0,$$

donc

$$\text{Cov}(f(U_{\pi^{(k)}i_1}), f(U_{\pi^{(k)}i_2})) \leq \text{Cov}(f(U_{\pi^{(k-1)}i_1}), f(U_{\pi^{(k-1)}i_2})).$$

On conclut que

$$\text{Var}(\hat{\mu}_k) \leq \text{Var}(\hat{\mu}_{k-1}),$$

ce qui signifie que la variance à la k^e étape de subdivision est inférieure à celle trouvée à l'étape précédente (étape $k-1$).

3.1.4. Applications de la méthode

Précédemment dans la démonstration on a trouvé

$$\text{Var}(\hat{\mu}_k) = \text{Var}(\hat{\mu}_{k-1})$$

$$-\frac{1}{n^2} \frac{1}{n(n-1)} E \left(\sum_{i_1 \neq i_2} \sum_j \left\{ \left(\mu_{i_1,j}(\pi^{(k-1)}) - \mu_{i_1}(\pi^{(k-1)}) \right) \left(\mu_{i_2,j}(\pi^{(k-1)}) - \mu_{i_2}(\pi^{(k-1)}) \right) \right\} \right),$$

c'est-à-dire que

$$\text{Var}(\hat{\mu}_k) = \text{Var}(\hat{\mu}_{k-1})$$

$$-\frac{1}{n^2} \left\{ \frac{1}{n(n-1)} \right\}^k \sum_{i_1 \neq l_1} \dots \sum_{i_k \neq l_k} \sum_j \left(\mu_{i_1 \dots i_{(k)}j} - \mu_{i_1 \dots i_{(k-1)}} \right) \left(\mu_{l_1 \dots l_{(k)}j} - \mu_{l_1 \dots l_{(k)}} \right).$$

3.1.4.1. Fonction identité

- Cas d'un tirage multinomial

Pour le cas multinomial on a

$$\text{Var}(\hat{\mu}_{(Mult)}) = \frac{\text{Var}(U)}{n} = \frac{1}{12n}.$$

- Tirage stratifié

A l'étape 0 ($k = 0$)

$$\text{Var}(\hat{\mu}_0) = \text{Var}(\hat{\mu}_{(Mult)}) - \frac{1}{n^2} \sum_i (\mu_i - \mu)^2. \quad (3.1.6)$$

On calcule μ_i , la moyenne pour chaque sous-intervalle I_i

$$\mu_i = \frac{2i-1}{n} \quad \text{et} \quad \mu = E(U) = \frac{1}{2}.$$

Or

$$\sum_i (\mu_i - \mu)^2 = \frac{(n+1)(n-1)}{12n},$$

car

$$\begin{aligned}
\sum_i (\mu_i - \mu)^2 &= \sum \left(\frac{i}{n} - \frac{1}{2n} - \frac{1}{2} \right)^2 \\
&= \frac{1}{n^2} \sum \left(i - \frac{n+1}{2} \right)^2.
\end{aligned}$$

Puisque

$$\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6} \quad \text{et} \quad \sum_{i=1}^n i = \frac{n(n+1)}{2},$$

on a

$$\begin{aligned}
\sum_i (\mu_i - \mu)^2 &= \frac{1}{n^2} \left(\sum_i i^2 - (n+1) \sum_i i + n(n+1) \right), \\
&= \frac{1}{n^2} \left(\frac{n(n+1)(2n+1)}{6} - 2(n+1) \frac{n(n+1)}{2} + \frac{n(n+1)^2}{4} \right), \\
&= \frac{(n+1)}{2n} \left(\frac{(2n+1)}{3} - (n+1) + \frac{n+1}{2} \right), \\
&= \frac{(n+1)(n-1)}{12n}, \\
&= \frac{n^2}{12n} \left(1 - \frac{1}{n^2} \right),
\end{aligned}$$

(3.1.6) devient

$$\begin{aligned}
 \text{Var}(\hat{\mu}_0) &= \text{Var}(\hat{\mu}_{(Mult)}) - \sum_i (\mu_i - \mu)^2 / n^2, \\
 &= \frac{1}{12n} - \frac{1}{12n} \left(1 - \frac{1}{n^2}\right), \\
 &= \frac{1}{12n} \frac{1}{n^2}.
 \end{aligned}$$

A l'étape 1 ($k = 1$)

$$\text{Var}(\hat{\mu}_1) = \text{Var}(\hat{\mu}_0) - \frac{1}{n^2} \sum_{ij} \left(\mu_{ij} - \mu_i \right)^2.$$

Par le même raisonnement on trouve que

$$\sum_{ij} \left(\mu_{ij} - \mu_i \right)^2 = \frac{1}{12n} \left(1 - \frac{1}{n^2}\right).$$

Donc

$$\begin{aligned}
 \text{Var}(\hat{\mu}_1) &= \left\{ \frac{1}{12n} - \frac{1}{12n} \left(1 - \frac{1}{n^2}\right) \right\} - \left\{ \frac{1}{12n^3} \left(1 - \frac{1}{n^2}\right) \right\}, \\
 &= \frac{1}{12n} \frac{1}{n^4}.
 \end{aligned}$$

Ainsi pour $k > 0$ quelconque on a

$$\text{Var}(\hat{\mu}_k) = \text{Var}(\hat{\mu}_{k-1}) - \frac{1}{n^{2k+1}} \left(1 - \frac{1}{n^2}\right),$$

d'où,

$$\begin{aligned}
 \text{Var}(\hat{\mu}_k) &= \text{Var}(\hat{\mu}_0) + \sum_{j=1}^k [\text{Var}(\hat{\mu}_j) - \text{Var}(\hat{\mu}_{j-1})], \\
 &= \frac{1}{12n^3} - \sum_{j=1}^k \frac{1}{n^{2j+1}} \left(1 - \frac{1}{n^2}\right), \\
 &= \frac{1}{12n^{k+3}}.
 \end{aligned}$$

3.1.4.2. *Modèle exponentiel*

Dans cette situation on se donne l'objectif d'estimer $E(\exp(tU))$ pour U de loi uniforme sur $(0, 1)$ et $t > 0$.

Dans le cas du **Tirage simple** on a

$$\mu = E(e^{tU}) = \int_0^1 e^{tU} du = \frac{1}{t}(e^t - 1)$$

et

$$\text{Var}(e^{tU}) = E(e^{2tU}) - (E(e^{tU}))^2.$$

Ainsi,

$$E(e^{2tU}) = \frac{1}{2t}(e^{2t} - 1) \quad \text{et}$$

$$\begin{aligned}
 \text{Var}(e^{tU}) &= \frac{1}{2t}(e^{2t} - 1) - \frac{1}{t^2}(e^t - 1)^2, \\
 &= \frac{1}{t}(e^t - 1) \left[\frac{(e^t + 1)}{2} - \frac{(e^t - 1)}{t} \right] \\
 &= \frac{1}{t}(e^t - 1) \left[\frac{(te^t + t - 2e^t + 2)}{2t} \right].
 \end{aligned}$$

Par conséquent,

$$\text{Var}(\hat{\mu}_{Mult}) = \frac{\text{Var}(e^{tU})}{n} = \frac{1}{2nt^2} \left(e^t - 1 \right) \left[\left(t + 2 \right) + \left(t - 2 \right) e^t \right].$$

A. Tirage stratifié à l'étape 0 ($k = 0$)

$$\text{Var}(\hat{\mu}_0) = \text{Var}(\hat{\mu}_{Mult}) - \frac{1}{n^2} \sum_i (\mu_i - \mu)^2,$$

$$\text{Var}(\hat{\mu}_0) = \frac{1}{2nt^2} \left(e^t - 1 \right) \left[\left(t + 2 \right) + \left(t - 2 \right) e^t \right] - \frac{1}{n^2} \sum_i (\mu_i - \mu)^2. \quad (3.1.7)$$

Soit

$$\begin{aligned} \mu_i &= n \int_{\frac{(i-1)}{n}}^{\frac{i}{n}} e^{tU} du, \\ &= \frac{n}{t} \left[e^{\frac{ti}{n}} - e^{\frac{t(i-1)}{n}} \right] = \frac{n}{t} \left[e^{\frac{t(i-1)}{n}} \left(e^{\frac{t}{n}} - 1 \right) \right], \end{aligned}$$

avec

$$\begin{aligned} \frac{1}{n} \sum_i \mu_i &= \frac{1}{n} \frac{n}{t} \left(e^{\frac{t}{n}} - 1 \right) \sum_i e^{t \frac{(i-1)}{n}} \\ &= \frac{1}{t} \left(e^t - 1 \right) = \mu, \end{aligned}$$

car

$$\sum_i e^{t \frac{(i-1)}{n}} = \frac{\left(e^t - 1 \right)}{\left(e^{\frac{t}{n}} - 1 \right)}.$$

Donc

$$\sum_i (\mu_i - \mu)^2 = \sum_i \mu_i^2 - n\mu^2.$$

Ainsi

$$\sum_i \mu_i^2 = \frac{n^2}{t^2} \left(e^{\frac{t}{n}} - 1 \right)^2 \sum_i e^{\frac{2t(i-1)}{n}}.$$

Or

$$\begin{aligned} \sum_i e^{\frac{2t(i-1)}{n}} &= \frac{e^{2t} - 1}{e^{\frac{2t}{n}} - 1}, \\ &= \frac{e^{2t} - 1}{\left(e^{\frac{t}{n}} + 1 \right) \left(e^{\frac{t}{n}} - 1 \right)}. \end{aligned}$$

Donc

$$\begin{aligned} \sum_i (\mu_i - \mu)^2 &= \frac{n^2}{t^2} \left(e^{\frac{t}{n}} - 1 \right)^2 \frac{e^{2t} - 1}{\left(e^{\frac{t}{n}} + 1 \right) \left(e^{\frac{t}{n}} - 1 \right)} - \frac{n}{t^2} \left(e^t - 1 \right)^2, \\ &= \frac{n^2}{t^2} \left(e^{\frac{t}{n}} - 1 \right) \frac{e^{2t} - 1}{\left(e^{\frac{t}{n}} + 1 \right)} - \frac{n}{t^2} \left(e^t - 1 \right)^2, \\ &= \frac{n}{t^2} \left(e^t - 1 \right) \left[\frac{n \left(e^t + 1 \right) \left(e^{\frac{t}{n}} - 1 \right)}{\left(e^{\frac{t}{n}} + 1 \right)} - \left(e^t - 1 \right) \right]. \end{aligned}$$

On a

$$\frac{1}{n^2} \sum_i (\mu_i - \mu)^2 = \frac{1}{nt^2} \left(e^t - 1 \right) \left[\frac{n \left(e^t + 1 \right) \left(e^{\frac{t}{n}} - 1 \right)}{\left(e^{\frac{t}{n}} + 1 \right)} - \left(e^t - 1 \right) \right].$$

Maintenant on développe l'expression (3.1.7)

$$\begin{aligned}
\text{Var}(\hat{\mu}_0) &= \frac{1}{2nt^2} (e^t - 1) \left[(t+2) + (t-2)e^t \right], \\
&- \frac{1}{nt^2} (e^t - 1) \left[\frac{n(e^t + 1)(e^{\frac{t}{n}} - 1)}{(e^{\frac{t}{n}} + 1)} - (e^t - 1) \right], \\
&= \frac{1}{nt^2} (e^t - 1) \left[\frac{(t+2) + (t-2)e^t}{2} - \left(\frac{n(e^t + 1)(e^{\frac{t}{n}} - 1)}{(e^{\frac{t}{n}} + 1)} - (e^t - 1) \right) \right], \\
&= \frac{1}{nt^2} (e^t - 1) \left[\frac{(t+2) + (t-2)e^t}{2} - \frac{2(e^t - 1)}{2} - \left(\frac{n(e^t + 1)(e^{\frac{t}{n}} - 1)}{(e^{\frac{t}{n}} + 1)} \right) \right].
\end{aligned}$$

Ceci donne

$$\text{Var}(\hat{\mu}_0) = \frac{1}{nt^2} (e^t - 1) \left[\frac{t(1 + e^t)}{2} - \left(\frac{n(e^t + 1)(e^{\frac{t}{n}} - 1)}{(e^{\frac{t}{n}} + 1)} \right) \right]. \quad (3.1.8)$$

Or

$$n \left(e^{\frac{t}{n}} - 1 \right) = \frac{\left(e^{\frac{t}{n}} - 1 \right)}{\frac{1}{n}} \xrightarrow{n \rightarrow \infty} t, \quad (3.1.9)$$

et

$$e^{\frac{t}{n}} + 1 \xrightarrow{n \rightarrow \infty} 2. \quad (3.1.10)$$

Donc

$$\left(\frac{n \left(e^t + 1 \right) \left(e^{\frac{t}{n}} - 1 \right)}{\left(e^{\frac{t}{n}} + 1 \right)} \right) \xrightarrow{n \rightarrow \infty} \frac{t \left(1 + e^t \right)}{2}. \quad (3.1.11)$$

Provisoirement on a

$$\text{Var}(\hat{\mu}_0) = \frac{1}{2nt^2} \left(e^t - 1 \right) \left\{ \frac{t \left(1 + e^t \right)}{2} - \frac{t \left(1 + e^t \right)}{2} \right\} \xrightarrow{n \rightarrow \infty} 0.$$

Donc la variance tend vers zéro lorsque n tend vers l'infini.

Pour obtenir la vitesse de convergence on doit considérer les développements suivants

$$\frac{n}{t} \left(e^{\frac{t}{n}} - 1 \right) = 1 + \frac{t}{2n} + \frac{t^2}{6n^2} + o\left(\frac{1}{n^2}\right), \quad (3.1.12)$$

$$e^{\frac{t}{n}} + 1 = 2 + \frac{t}{n} + \frac{t^2}{2n^2} + o\left(\frac{1}{n^2}\right). \quad (3.1.13)$$

Or d'après (3.1.12) et (3.1.13), (3.1.8) devient

$$\begin{aligned}
\text{Var}(\hat{\mu}_0) &= \frac{1}{nt^2} (e^t - 1) (1 + e^t) t \left[\frac{1}{2} - \frac{n \left(e^{\frac{t}{n}} - 1 \right)}{t \left(e^{\frac{t}{n}} + 1 \right)} \right], \\
&= \frac{1}{nt^2} \frac{(e^t - 1)(1 + e^t)t}{2 \left(e^{\frac{t}{n}} + 1 \right)} \left(\left(e^{\frac{t}{n}} + 1 \right) - \frac{2n}{t} \left(e^{\frac{t}{n}} - 1 \right) \right), \\
&= \frac{1}{nt^2} \frac{(e^t - 1)(1 + e^t)t}{2 \left(e^{\frac{t}{n}} + 1 \right)} \left(2 + \frac{t}{n} + \frac{t^2}{2n^2} + o\left(\frac{1}{n^2}\right) - 2 \left(1 + \frac{t}{2n} + \frac{t^2}{6n^2} + o\left(\frac{1}{n^2}\right) \right) \right), \\
&= \frac{1}{nt} \frac{(e^t - 1)(1 + e^t)}{4} \left[\left(\frac{t^2}{2n^2} + o\left(\frac{1}{n^2}\right) \right) - \left(\frac{t^2}{3n^2} + o\left(\frac{1}{n^2}\right) \right) \right].
\end{aligned}$$

Donc

$$\begin{aligned}
\text{Var}(\hat{\mu}_0) &= \frac{1}{nt} \frac{(e^t - 1)(1 + e^t)}{4} \left(\frac{t^2}{6n^2} + o\left(\frac{1}{n^2}\right) \right), \\
&= \frac{t}{24n^3} (e^t - 1)(1 + e^t) + o\left(\frac{1}{n^3}\right).
\end{aligned}$$

Alors on a

$$n^3 \text{Var}(\hat{\mu}_0) \xrightarrow{n \rightarrow \infty} \frac{t(e^t - 1)(1 + e^t)}{24}.$$

On conclut que la variance à l'étape 0 est de l'ordre de n^{-3} .

B. Tirage stratifié à l'étape 1 ($k = 1$)

Dans ce cas-ci on a :

$$\text{Var}(\hat{\mu}_1) = \text{Var}(\mu_0) - \frac{1}{n^2} \frac{1}{n(n-1)} \sum_l \sum_{i \neq j} \sum \left(\mu_{il} - \mu_i \right) \left(\mu_{jl} - \mu_j \right) r, \quad (3.1.14)$$

où $\text{Var}(\mu_0)$ est la variance pour $k = 0$ calculée précédemment et

$$\sum_l \sum_{i \neq j} \sum \left(\mu_{il} - \mu_i \right) \left(\mu_{jl} - \mu_j \right)$$

$$= \sum_l \sum_i \sum_j \left(\mu_{il} - \mu_i \right) \left(\mu_{jl} - \mu_j \right) - \sum_l \sum_i \left(\mu_{il} - \mu_i \right)^2, \quad (3.1.15)$$

$$= \sum_l \left(\sum_i \left(\mu_{il} - \mu_i \right) \right)^2 - \sum_l \sum_i \left(\mu_{il} - \mu_i \right)^2. \quad (3.1.16)$$

Calculons les deux expressions de (3.1.16) séparément.

-Commençons par $\sum_l \left(\sum_i \left(\mu_{il} - \mu_i \right) \right)^2$

On commence par définir

$$\begin{aligned} \mu_{il} &= n^2 \int_{\frac{i-1}{n} + \frac{l-1}{n^2}}^{\frac{i+1}{n} + \frac{l}{n^2}} e^{tU} du, \\ &= \frac{n^2}{t} e^{t(\frac{i-1}{n} + \frac{l-1}{n^2})} - e^{t(\frac{i-1}{n} + \frac{l-1}{n^2})}, \\ &= \frac{n^2}{t} e^{t(\frac{i-1}{n} + \frac{l-1}{n^2})} \left(e^{\frac{t}{n^2}} - 1 \right), \end{aligned}$$

et

$$\begin{aligned}
 \sum_i \mu_{il} &= \sum_i \frac{n^2}{t} e^{t(\frac{i-1}{n} + \frac{l-1}{n^2})} \left(e^{\frac{t}{n^2}} - 1 \right), \\
 &= \frac{n^2}{t} e^{t(\frac{l-1}{n^2})} \left(e^{\frac{t}{n^2}} - 1 \right) \sum_i e^{t(\frac{i-1}{n})}, \\
 &= \frac{n^2}{t} e^{t(\frac{l-1}{n^2})} \left(e^{\frac{t}{n^2}} - 1 \right) \frac{(e^t - 1)}{\left(e^{\frac{t}{n}} - 1 \right)},
 \end{aligned}$$

car

$$\sum_i e^{t(\frac{i-1}{n})} = \frac{e^t - 1}{e^{\frac{t}{n}} - 1},$$

et

$$\begin{aligned}
 \sum_i \mu_i &= \frac{n}{t} \left(e^{\frac{t}{n}} - 1 \right) \sum e^{t(\frac{i-1}{n})}, \\
 &= \frac{n}{t} \left(e^{\frac{t}{n}} - 1 \right) \frac{e^t - 1}{e^{\frac{t}{n}} - 1}, \\
 &= \frac{n}{t} (e^t - 1).
 \end{aligned}$$

On a donc

$$\begin{aligned}
 \sum_i \mu_{il} - \sum_i \mu_i &= \frac{n^2}{t} e^{t(\frac{l-1}{n^2})} \left(e^{\frac{t}{n^2}} - 1 \right) \frac{(e^t - 1)}{\left(e^{\frac{t}{n}} - 1 \right)} - \frac{n}{t} (e^t - 1), \\
 &= \frac{n}{t} (e^t - 1) \left[\frac{n \left(e^{\frac{t}{n^2}} - 1 \right) e^{t(\frac{l-1}{n^2})}}{\left(e^{\frac{t}{n}} - 1 \right)} - 1 \right],
 \end{aligned}$$

$$\begin{aligned}
\left(\sum_{\mathbf{i}} \mu_{\mathbf{il}} - \sum_{\mathbf{i}} \mu_{\mathbf{i}} \right)^2 &= \left\{ \frac{n}{t} (e^t - 1) \left[\frac{n \left(e^{\frac{t}{n^2}} - 1 \right) e^{t \left(\frac{l-1}{n^2} \right)}}{\left(e^{\frac{t}{n}} - 1 \right)} - 1 \right] \right\}^2, \\
&= \frac{n^2}{t^2} (e^t - 1)^2 \left\{ \frac{n^2 \left(e^{\frac{t}{n^2}} - 1 \right)^2 e^{\frac{2t(l-1)}{n^2}}}{\left(e^{\frac{t}{n}} - 1 \right)^2} - \frac{2n \left(e^{\frac{t}{n^2}} - 1 \right) e^{\frac{t(l-1)}{n}}}{\left(e^{\frac{t}{n}} - 1 \right)} + 1 \right\}.
\end{aligned}$$

Et le calcul de $\sum_l \left(\sum_i \mu_{il} - \sum_i \mu_i \right)^2$ nous donne

$$= \frac{n^2}{t^2} (e^t - 1)^2 \left\{ \frac{n^2 \left(e^{\frac{t}{n^2}} - 1 \right)^2 \sum_l \left(e^{\frac{2t(l-1)}{n^2}} \right)}{\left(e^{\frac{t}{n}} - 1 \right)^2} - \frac{2n \left(e^{\frac{t}{n^2}} - 1 \right) \sum_l \left(e^{\frac{t(l-1)}{n^2}} \right)}{\left(e^{\frac{t}{n}} - 1 \right)} + n \right\}.$$

Or

$$\sum_l \left(e^{\frac{2t(l-1)}{n^2}} \right) = \frac{\left(e^{\frac{2t}{n}} - 1 \right)}{\left(e^{\frac{2t}{n^2}} - 1 \right)},$$

et

$$\sum_l \left(e^{\frac{t(l-1)}{n^2}} \right) = \frac{\left(e^{\frac{t}{n}} - 1 \right)}{\left(e^{\frac{t}{n^2}} - 1 \right)}.$$

Donc

$$\begin{aligned}
&= \frac{n^2}{t^2} \left(e^t - 1 \right)^2 \left\{ \frac{n^2 \left(e^{\frac{t}{n^2}} - 1 \right)^2 \left(e^{\frac{2t}{n}} - 1 \right)}{\left(e^{\frac{t}{n}} - 1 \right)^2 \left(e^{\frac{2t}{n^2}} - 1 \right)} - \frac{2n \left(e^{\frac{t}{n^2}} - 1 \right) \left(e^{\frac{t}{n}} - 1 \right)}{\left(e^{\frac{t}{n}} - 1 \right) \left(e^{\frac{t}{n^2}} - 1 \right)} + n \right\}, \\
&= \frac{n^2}{t^2} \left(e^t - 1 \right)^2 \left\{ \frac{n^2 \left(e^{\frac{t}{n^2}} - 1 \right) \left(e^{\frac{t}{n}} + 1 \right)}{\left(e^{\frac{t}{n}} - 1 \right) \left(e^{\frac{t}{n^2}} + 1 \right)} - 2n + n \right\}, \\
&= \frac{n^2}{t^2} \left(e^t - 1 \right)^2 \left\{ \frac{n^2 \left(e^{\frac{t}{n^2}} - 1 \right) \left(e^{\frac{t}{n}} + 1 \right)}{\left(e^{\frac{t}{n}} - 1 \right) \left(e^{\frac{t}{n^2}} + 1 \right)} - n \right\},
\end{aligned}$$

car

$$\left(e^{\frac{2t}{n^2}} - 1 \right) = \left(e^{\frac{t}{n^2}} - 1 \right) \left(e^{\frac{t}{n^2}} + 1 \right)$$

et

$$\left(e^{\frac{2t}{n}} - 1 \right) = \left(e^{\frac{t}{n}} - 1 \right) \left(e^{\frac{t}{n}} + 1 \right),$$

$$\sum_l \left(\sum_i \mu_{il} - \sum_i \mu_i \right)^2 = \frac{n^3}{t^2} \left(e^t - 1 \right)^2 \left\{ \frac{n \left(e^{\frac{t}{n^2}} - 1 \right) \left(e^{\frac{t}{n}} + 1 \right)}{\left(e^{\frac{t}{n}} - 1 \right) \left(e^{\frac{t}{n^2}} + 1 \right)} - 1 \right\}.$$

Puisque

$$\frac{n^2}{t} \left(e^{\frac{t}{n^2}} - 1 \right) = \frac{\left(e^{\frac{t}{n^2}} - 1 \right)}{\frac{t}{n^2}} \xrightarrow{n \rightarrow \infty} 1, \quad (3.1.17)$$

et également

$$\frac{n}{t} \left(e^{\frac{t}{n}} - 1 \right) = \frac{\left(e^{\frac{t}{n}} - 1 \right)}{\frac{t}{n}} \xrightarrow{n \rightarrow \infty} 1, \quad (3.1.18)$$

on a

$$\frac{n^2 \left(e^{\frac{t}{n^2}} - 1 \right) \left(e^{\frac{t}{n}} + 1 \right)}{n \left(e^{\frac{t}{n}} - 1 \right) \left(e^{\frac{t}{n^2}} + 1 \right)} \xrightarrow{n \rightarrow \infty} 1,$$

et

$$\frac{n^2 \left(e^{\frac{t}{n^2}} - 1 \right) \left(e^{\frac{t}{n}} + 1 \right)}{n \left(e^{\frac{t}{n}} - 1 \right) \left(e^{\frac{t}{n^2}} + 1 \right)} - 1 \xrightarrow{n \rightarrow \infty} 0.$$

Ainsi

$$\sum_l \left(\sum_i \mu_{il} - \sum_i \mu_i \right)^2 \xrightarrow{n \rightarrow \infty} 0.$$

-Calculons $\sum_l \sum_i \left(\mu_{il} - \mu_i \right)^2$.

On a

$$\sum_l \sum_i \left(\mu_{il} - \mu_i \right)^2 = \sum_i \sum_l \left(\mu_{il} - \mu_i \right)^2 = \sum_i \left(\sum_l \mu_{il}^2 - n \mu_i^2 \right),$$

car

$$\begin{aligned} \frac{1}{n} \sum_l \mu_{il} &= \frac{1}{n} \frac{n^2}{t} \left(e^{\frac{t}{n^2}} - 1 \right) e^{\frac{t(i-1)}{n}} \sum_l e^{\frac{(l-1)t}{n^2}}, \\ &= \frac{n}{t} \frac{\left(e^{\frac{t}{n^2}} - 1 \right) e^{\frac{t(i-1)}{n}} \left(e^{\frac{t}{n}} - 1 \right)}{\left(e^{\frac{t}{n^2}} - 1 \right)}, \\ &= \frac{n}{t} e^{\frac{t(i-1)}{n}} \left(e^{\frac{t}{n}} - 1 \right) = \mu_i. \end{aligned}$$

Commençons par calculer $\sum_l \mu_{il}^2$.

On a

$$\begin{aligned}
 \sum_l \mu_{il}^2 &= \sum_l \left(\frac{n^2}{t} \left(e^{\frac{t}{n^2}} - 1 \right) e^{\frac{t(i-1)}{n}} e^{\frac{t(l-1)}{n^2}} \right)^2, \\
 &= \sum_l \left(\frac{n^4}{t^2} \left(e^{\frac{t}{n^2}} - 1 \right)^2 e^{\frac{2(i-1)}{n}} e^{\frac{2(l-1)}{n^2}} \right), \\
 &= \frac{n^4}{t^2} \left(e^{\frac{t}{n^2}} - 1 \right)^2 e^{\frac{2t(i-1)}{n}} \sum_l e^{\frac{2t(l-1)}{n^2}}, \\
 &= \frac{n^4 \left(e^{\frac{t}{n^2}} - 1 \right)^2 e^{\frac{2t(i-1)}{n}} (e^{\frac{2t}{n}} - 1)}{t^2 (e^{\frac{2t}{n^2}} - 1)},
 \end{aligned}$$

et

$$\begin{aligned}
 \sum_i \left(\sum_l \mu_{il}^2 \right) &= \sum_i \frac{n^4 \left(e^{\frac{t}{n^2}} - 1 \right)^2 e^{\frac{2t(i-1)}{n}} (e^{\frac{2t}{n}} - 1)}{t^2 (e^{\frac{2t}{n^2}} - 1)}, \\
 &= \frac{n^4 \left(e^{\frac{t}{n^2}} - 1 \right)^2 (e^{\frac{2t}{n}} - 1)}{t^2 (e^{\frac{2t}{n^2}} - 1)} \sum_i e^{\frac{2t(i-1)}{n}}, \\
 &= \frac{n^4 \left(e^{\frac{t}{n^2}} - 1 \right)^2 (e^{\frac{2t}{n}} - 1) (e^{2t} - 1)}{t^2 (e^{\frac{t}{n^2}} + 1) (e^{\frac{2t}{n}} - 1)}.
 \end{aligned}$$

Donc

$$\sum_i \left(\sum_l \mu_{il}^2 \right) = \frac{n^4 \left(e^{\frac{t}{n^2}} - 1 \right)^2 (e^{2t} - 1)}{t^2 (e^{\frac{t}{n^2}} + 1)}.$$

Calcul de $\sum_i \mu_i^2$.

On a

$$\begin{aligned}
 \sum_i \mu_i^2 &= \left(\frac{n}{t} e^{\frac{t(i-1)}{n}} \left(e^{\frac{t}{n}} - 1 \right) \right)^2, \\
 &= \sum_i \left(\frac{n^2}{t^2} e^{\frac{2t(i-1)}{n}} \left(e^{\frac{t}{n}} - 1 \right)^2 \right), \\
 &= \frac{n^2}{t^2} \left(e^{\frac{t}{n}} - 1 \right)^2 \sum_i e^{\frac{2t(i-1)}{n}}, \\
 &= \frac{n^2}{t^2} \frac{\left(e^{\frac{t}{n}} - 1 \right)^2 \left(e^{2t} - 1 \right)}{\left(e^{\frac{2t}{n}} - 1 \right)}, \\
 &= \frac{n^2}{t^2} \frac{\left(e^{\frac{t}{n}} - 1 \right) \left(e^{2t} - 1 \right)}{\left(e^{\frac{t}{n}} + 1 \right)},
 \end{aligned}$$

et

$$n \sum_i \mu_i^2 = \frac{n^3}{t^2} \frac{\left(e^{\frac{t}{n}} - 1 \right) \left(e^{2t} - 1 \right)}{\left(e^{\frac{t}{n}} + 1 \right)}.$$

Alors

$$\begin{aligned}
 \sum_l \sum_i \left(\mu_{il} - \mu_i \right)^2 &= \sum_i \left(\sum_l \mu_{il}^2 - n \mu_i^2 \right) \\
 &= \left\{ \left(\frac{n^4}{t^2} \frac{\left(e^{\frac{t}{n^2}} - 1 \right) \left(e^{2t} - 1 \right)}{\left(e^{\frac{t}{n^2}} + 1 \right)} \right) - \left(\frac{n^3}{t^2} \frac{\left(e^{\frac{t}{n}} - 1 \right) \left(e^{2t} - 1 \right)}{\left(e^{\frac{t}{n}} + 1 \right)} \right) \right\}.
 \end{aligned}$$

Donc

$$\sum_i \left(\sum_l \mu_{il}^2 - n \mu_i^2 \right) = \frac{n^2}{t^2} (e^{2t} - 1) \left\{ \left(\frac{n^2 \left(e^{\frac{t}{n^2}} - 1 \right)}{\left(e^{\frac{t}{n^2}} + 1 \right)} \right) - \left(\frac{n \left(e^{\frac{t}{n}} - 1 \right)}{\left(e^{\frac{t}{n}} + 1 \right)} \right) \right\}. \quad (3.1.19)$$

On remarque que d'après (3.1.17) et (3.1.18)

$$\left\{ \frac{n^2 \left(e^{\frac{t}{n^2}} - 1 \right)}{\left(e^{\frac{t}{n^2}} + 1 \right)} \right\} \xrightarrow{n \rightarrow \infty} \frac{t}{2},$$

et

$$\left\{ \frac{n \left(e^{\frac{t}{n}} - 1 \right)}{\left(e^{\frac{t}{n}} + 1 \right)} \right\} \xrightarrow{n \rightarrow \infty} \frac{t}{2}.$$

Ainsi donc

$$\sum_l \sum_i \left(\mu_{il} - \mu_i \right)^2 = \sum_i \left(\sum_l \mu_{il}^2 - n \mu_i^2 \right) \xrightarrow{n \rightarrow \infty} 0.$$

Pour l'expression (3.1.19) on a trouvé

$$\begin{aligned}
\sum_l \left(\sum_i \mu_{il} - \sum_i \mu_i \right)^2 &= \frac{n^3}{t^2} (e^t - 1)^2 \left\{ \frac{n \left(e^{\frac{t}{n^2}} - 1 \right) \left(e^{\frac{t}{n}} + 1 \right)}{\left(e^{\frac{t}{n}} - 1 \right) \left(e^{\frac{t}{n^2}} + 1 \right)} - 1 \right\}, \\
&= \frac{n^2 (e^t - 1)^2}{t \left(e^{\frac{t}{n}} - 1 \right) \left(e^{\frac{t}{n^2}} + 1 \right)} \left[\frac{n^2}{t} \left(e^{\frac{t}{n^2}} - 1 \right) \left(e^{\frac{t}{n}} + 1 \right), \right. \\
&\quad \left. - \frac{n}{t} \left(e^{\frac{t}{n}} - 1 \right) \left(e^{\frac{t}{n^2}} + 1 \right) \right].
\end{aligned}$$

D'après (3.1.12), (3.1.13) et

$$\frac{n^2}{t} \left(e^{\frac{t}{n^2}} - 1 \right) = 1 + \frac{t}{2n^2} + \frac{t^2}{6n^4} + o\left(\frac{1}{n^4}\right), \quad (3.1.20)$$

$$\left(e^{\frac{t}{n^2}} + 1 \right) = 2 + \frac{t}{n^2} + \frac{t^2}{2n^4} + o\left(\frac{1}{n^4}\right). \quad (3.1.21)$$

Donc on a

$$\begin{aligned}
\sum_l \left(\sum_i \mu_{il} - \sum_i \mu_i \right)^2 &= \frac{n^2 (e^t - 1)^2}{t \left(e^{\frac{t}{n}} - 1 \right) \left(e^{\frac{t}{n^2}} + 1 \right)}, \\
&\quad \left[\left(1 + \frac{t}{2n^2} + \frac{t^2}{6n^4} + o\left(\frac{1}{n^4}\right) \right) \left(2 + \frac{t}{n} + \frac{t^2}{2n^2} + o\left(\frac{1}{n^2}\right) \right), \right. \\
&\quad \left. - \left(2 + \frac{t}{n^2} + \frac{t}{2n^4} + o\left(\frac{1}{n^4}\right) \right) \left(1 + \frac{t}{2n} + \frac{t^2}{6n^2} + o\left(\frac{1}{n^2}\right) \right) \right].
\end{aligned}$$

Alors

$$\begin{aligned} \sum_l \left(\sum_i \mu_{il} - \sum_i \mu_i \right)^2 &= \frac{n^2 \left(e^t - 1 \right)^2}{t \left(e^{\frac{t}{n}} - 1 \right) \left(e^{\frac{t}{n^2}} + 1 \right)} \left[\frac{t^2}{2n^2} - \frac{2t^2}{6n^2} + o\left(\frac{1}{n^2}\right) \right], \\ &= \frac{n^2 \left(e^t - 1 \right)^2}{t \left(e^{\frac{t}{n}} - 1 \right) \left(e^{\frac{t}{n^2}} + 1 \right)} \left[\frac{t^2}{6n^2} + o\left(\frac{1}{n^2}\right) \right], \end{aligned}$$

d'où

$$\sum_l \left(\sum_i \mu_{il} - \sum_i \mu_i \right)^2 = \frac{n^3}{12} \left(e^t - 1 \right)^2 + o(n^3). \quad (3.1.22)$$

Pour la deuxième expression $\sum_l \sum_i \left(\mu_{il} - \mu_i \right)^2$

$$\begin{aligned} \sum_l \sum_i \left(\mu_{il} - \mu_i \right)^2 &= \frac{n^2}{t^2} \left(e^{2t} - 1 \right) \left\{ \left(\frac{n^2 \left(e^{\frac{t}{n^2}} - 1 \right)}{\left(e^{\frac{t}{n^2}} + 1 \right)} \right) - \left(\frac{n \left(e^{\frac{t}{n}} - 1 \right)}{\left(e^{\frac{t}{n}} + 1 \right)} \right) \right\}, \\ &= \frac{n^2 \left(e^{2t} - 1 \right)}{t^2 \left(e^{\frac{t}{n^2}} + 1 \right) \left(e^{\frac{t}{n}} + 1 \right)} \left\{ n^2 \left(e^{\frac{t}{n^2}} - 1 \right) \left(e^{\frac{t}{n}} + 1 \right), \right. \\ &\quad \left. - n \left(e^{\frac{t}{n}} - 1 \right) \left(e^{\frac{t}{n^2}} + 1 \right) \right\}. \end{aligned}$$

D'après (3.1.12), (3.1.13), (3.1.20) et (3.1.21) on a

$$\begin{aligned}
& \sum_l \sum_i \left(\mu_{il} - \mu_i \right)^2 \\
&= \frac{n^2 \left(e^{2t} - 1 \right)}{t \left(e^{\frac{t}{n^2}} + 1 \right) \left(e^{\frac{t}{n}} + 1 \right)} \left[\left(1 + \frac{t}{2n^2} + \frac{t^2}{6n^4} + o\left(\frac{1}{n^4}\right) \right) \left(2 + \frac{t}{n} + \frac{t^2}{2n^2} + o\left(\frac{1}{n^2}\right) \right) \right. \\
&\quad \left. - \left(2 + \frac{t}{n^2} + \frac{t^2}{2n^4} + o\left(\frac{1}{n^4}\right) \right) \left(1 + \frac{t}{2n} + \frac{t^2}{6n^2} + o\left(\frac{1}{n^2}\right) \right) \right].
\end{aligned}$$

Donc on a

$$\sum_l \sum_i \left(\mu_{il} - \mu_i \right)^2 = \frac{n^2 \left(e^{2t} - 1 \right)}{t} \left[\frac{t^2}{24n^2} + o\left(\frac{1}{n^2}\right) \right],$$

Ainsi

$$\begin{aligned}
& \frac{1}{n^2} \frac{1}{n(n-1)} \sum_l \sum_{i \neq j} \sum \left(\mu_{il} - \mu_i \right) \left(\mu_{jl} - \mu_j \right), \\
&= \frac{1}{n^2} \frac{1}{n(n-1)} \left[\left(\frac{n^3 \left(e^t - 1 \right)^2}{12t^2} \left(\frac{t^2}{n^2} + o\left(\frac{1}{n^2}\right) \right) \right) - \left(\frac{n^2 \left(e^{2t} - 1 \right)}{24t} \left(\frac{t^2}{n^2} + o\left(\frac{1}{n^2}\right) \right) \right) \right], \\
&= \frac{1}{n^2} \frac{1}{n(n-1)} \left[\frac{2n^3 \left(e^t - 1 \right)^2 - n^2 \left(e^{2t} - 1 \right) t}{24n^2} \right] + o(n^{-3}), \\
&= \frac{1}{24n^3} \left(e^t - 1 \right) \left[\frac{2n \left(e^t - 1 \right) - t \left(e^t + 1 \right)}{n-1} \right] + o(n^{-3}).
\end{aligned}$$

Donc l'expression (3.1.14) devient

$$\text{Var}(\hat{\mu}_1) = \text{Var}(\hat{\mu}_0) - \frac{1}{24n^3} \left(e^t - 1 \right) \left[\frac{2n \left(e^t - 1 \right) - t \left(e^t + 1 \right)}{n - 1} \right] + o(n^{-3}).$$

Or

$$\text{Var}(\hat{\mu}_0) = \frac{t}{24n^3} \left(e^t - 1 \right) \left(1 + e^t \right) + o(n^{-3}).$$

Donc

$$\begin{aligned} \text{Var}(\hat{\mu}_1) &= \frac{t}{24n^3} \left(e^t - 1 \right) \left(1 + e^t \right) - \frac{1}{24n^3} \left(e^t - 1 \right) \left[\frac{2n \left(e^t - 1 \right) - t \left(e^t + 1 \right)}{n - 1} \right] + o(n^{-3}), \\ &= \frac{(e^t - 1)}{24n^3} \left[\left(1 + e^t \right) t - \left(\frac{2n \left(e^t - 1 \right) - t \left(e^t + 1 \right)}{n - 1} \right) \right] + o(n^{-3}). \end{aligned}$$

On a

$$n^3 \text{Var}(\hat{\mu}_1) = \frac{(e^t - 1)}{24} \left[\left(1 + e^t \right) t - \left(\frac{2n \left(e^t - 1 \right) - t \left(e^t + 1 \right)}{n - 1} \right) \right] + o(1).$$

Or

$$\frac{2n \left(e^t - 1 \right)}{n - 1} \xrightarrow{n \rightarrow \infty} 2(e^t - 1),$$

et

$$\frac{(e^t - 1)}{n - 1} \xrightarrow{n \rightarrow \infty} 0.$$

Donc

$$n^3 \text{Var}(\hat{\mu}_1) \xrightarrow{n \rightarrow \infty} \frac{1}{24} (e^t - 1) \left[(te^t + t) - 2(e^t - 1) \right].$$

On trouve finalement que

$$n^3 \text{Var}(\mu_1) \xrightarrow{n \rightarrow \infty} \frac{1}{24} (e^t - 1) \left[(t - 2)e^t + (2 + t) \right]$$

donc la variance à l'étape 1 est de nouveau de l'ordre de n^{-3} , comme à l'étape 0.

3.1.5. Résultats de simulation

Dans cette partie on va présenter les courbes des résultats de simulation des variances obtenues pour estimer $E(X)$ avec le modèle exponentiel, $E(X)$, $E(\exp(X))$, $E(\cos(tX))$ avec le modèle uniforme où $t \in \mathbb{R}$ pour $k = 1, 2, 3$.

3.1.5.1. Modèle exponentiel

Si on considère $N = 3000$, le nombre de simulations, pour $m = 40$ le nombre de sections pour un échantillon de taille $n = 200$, les résultats qu'on obtient sont :

$$\text{Var}(\hat{\mu}_1) = 0,0001213569, \text{Var}(\hat{\mu}_2) = 0,0001247872 \text{Var}(\hat{\mu}_3) = 0.00006196942.$$

Donc

$$\text{Var}(\hat{\mu}_1) < \text{Var}(\hat{\mu}_2).$$

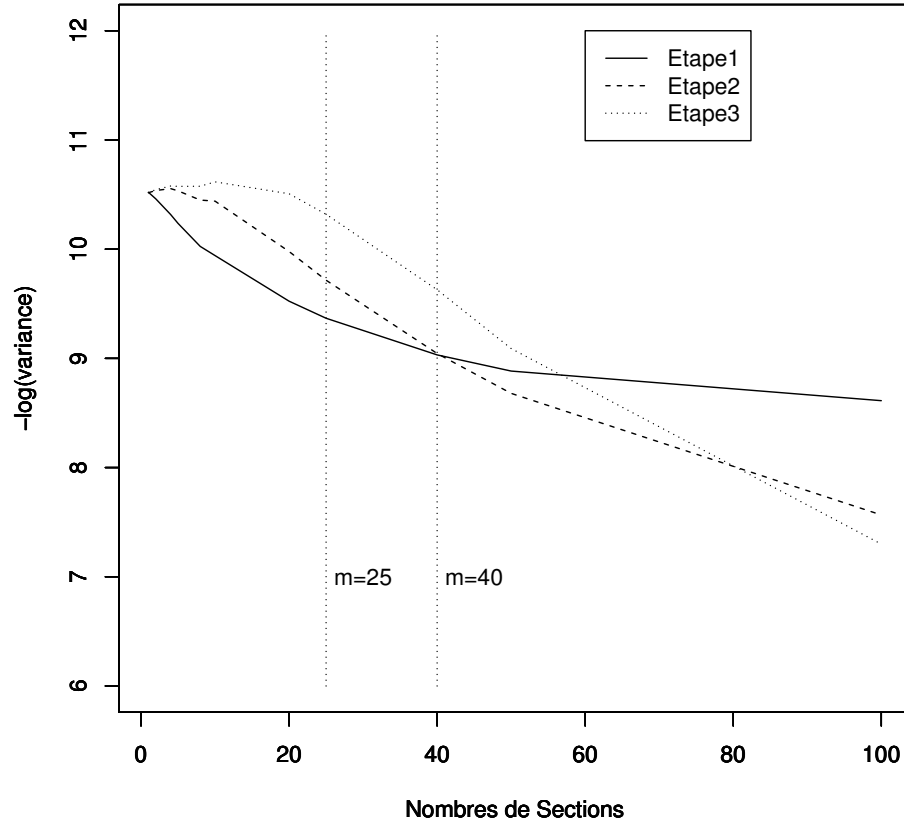


FIGURE 3.1. Variances : $E(X)$ par le modèle exponentiel pour $n = 200$

Mais si on augmente le nombre de simulations à $N = 50000$, on trouve pour $m = 40$ que

$$\text{Var}(\hat{\mu}_1) = 0,0001195521, \quad \text{Var}(\hat{\mu}_2) = 0,0001180394, \quad \text{Var}(\hat{\mu}_3) = 0,0000658164.$$

Donc

$$\text{Var}(\hat{\mu}_3) < \text{Var}(\hat{\mu}_2) < \text{Var}(\hat{\mu}_1).$$

Ce qui nous permet de conclure que si on continue à augmenter N , le nombre de simulations, on finira par trouver que

$$\text{Var}(\hat{\mu}_k) \leq \text{Var}(\hat{\mu}_{k-1}), \forall k > 1$$

où k représente la k^e étape de subdivision.

En d'autres termes on conclut que les trois courbes ne se croisent pas.

3.1.5.2. *Espérance de X pour le modèle uniforme*

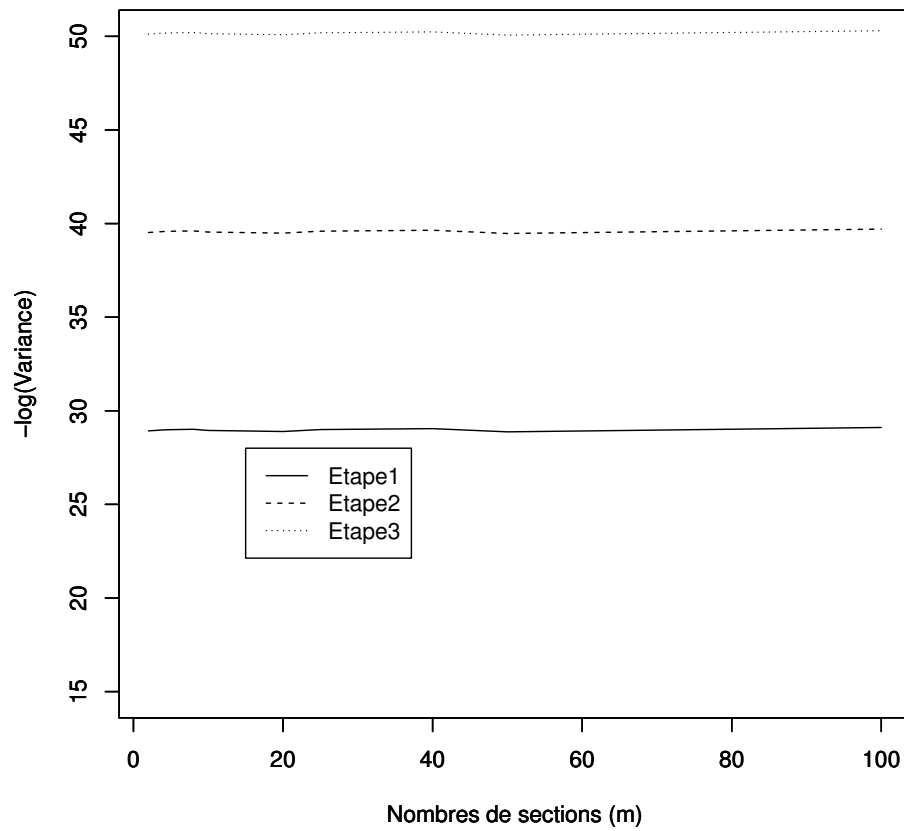


FIGURE 3.2. Variances : $E(X)$ pour le modèle uniforme pour $n = 200$

Sur le graphique toutes les courbes sont presque parallèles. A l'aide des formules on trouve par simple calcul que pour un échantillon de taille $n = 200$ on

$$\frac{\text{Var}(\hat{\mu}_1)}{\text{Var}(\hat{\mu}_2)} = \frac{1}{n^2} = \frac{1}{(200)^2} = 2.5 \times 10^{-5}$$

Ainsi le rapport entre les étapes qui se suivent est toujours le même et il est de l'ordre de 2.5×10^{-5} .

3.1.5.3. *Espérance de e^X pour le modèle uniforme*

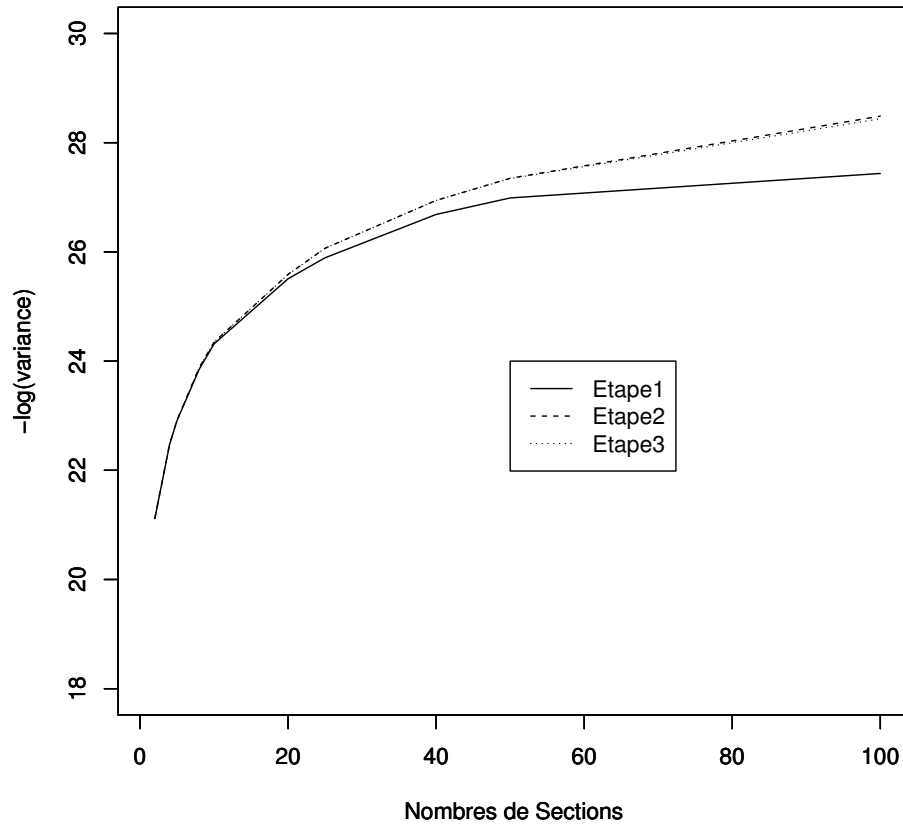


FIGURE 3.3. Variances : $E(e^X)$ pour le modèle uniforme avec $n = 200$

Sur le graphique on remarque que les courbes des variances à partir de la deuxième étape sont confondues donc pour ce modèle on va se limiter sur la première étape ($k = 1$). Le gain de réduction de la variance entre l'étape 1 et d'autres étapes pour $m = 100$ est

$$\frac{\text{Var}(\hat{\mu}_2)}{\text{Var}(\hat{\mu}_1)} = 2,858567.$$

3.1.5.4. *Espérance de $\cos(tX)$ pour le modèle uniforme*

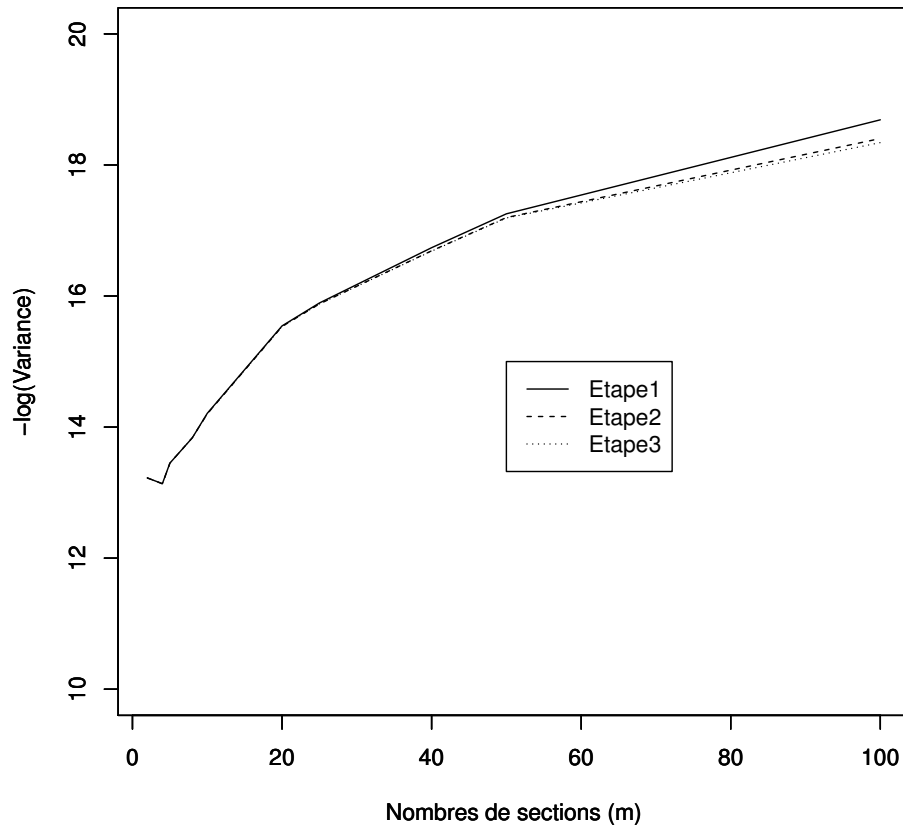


FIGURE 3.4. Variances : $E(\cos(tX))$ pour le modèle uniforme avec $n = 200, t = 20$

Dans cette situation on constate qu'au fur et à mesure que m augmente $\text{Var}(\hat{\mu}_i) < \text{Var}(\hat{\mu}_{i+1})$, $i = 1, 2$, d'où la contradiction.

Donc on peut donc conclure que pour la fonction $\cos(tX)$, la méthode proposée ne marche pas pour des grandes valeurs de t et de m .

3.1.6. Discussions

3.1.6.1. Cas où X n'est pas de loi uniforme sur $(0, 1)$

On cherche à estimer

$$I = E(f(x)) = \int_0^1 f(x)dx$$

où X n'est pas de loi uniforme sur $(0, 1)$ et f est une fonction croissante.

Donc si la variable X est générée par la méthode de transformation inverse, alors

$$X = F^{-1}(U) \quad \text{où } U \text{ est de loi uniforme sur } (0, 1).$$

On a

$$E(f(x)) = E[f \circ F^{-1}(U)].$$

La fonction $f \circ F^{-1}$ est croissante car f est une fonction croissante. Dans cette situation il faut être capable de trouver $F^{-1}(U)$.

Si on est dans l'impossibilité de connaître l'inverse de la fonction de répartition de F on utilise la méthode de l'échantillonnage pondéré pour trouver la densité instrumentale g facile à simuler et qui respecte toutes les propriétés déjà citées dans la section (2.3).

Donc on a

$$E[f(x)] = E\left[\frac{f}{g}(Y)\right] = E\left[\frac{f}{g} \circ G^{-1}(U)\right]$$

où g est la densité instrumentale et G^{-1} l'inverse de la fonction de répartition de G , U est de loi uniforme sur $(0, 1)$ et la variable Y suit la loi de densité g .

3.1.6.2. *Cas où la fonction f n'est pas croissante*

Dans cette situation on va jouer sur le nombre m , c'est à dire qu'on va diviser la fonction f en parties localement monotones et on va appliquer notre méthode à toutes parties où f est monotone et on suppose que l'influence de la partie où la fonction n'est pas monotone est négligeable pour la réduction de la variance car elle représente j/m pour $m \gg j$ où j est le nombre d'intervalles qui contiennent les points où la fonction n'est pas monotone.

CONCLUSION

Ce document est divisé en trois parties dont la première partie décrit trois différentes méthodes de génération des variables aléatoires, la deuxième partie traite en profondeur les aspects théoriques et pratiques de réduction de la variance : Méthode d'échantillonnage stratifié, méthode d'échantillonnage pondéré, méthode de variables de contrôle et méthode de variables antithétiques.

Dans ce mémoire on a beaucoup insisté sur la technique des variables antithétiques qui consiste à induire des corrélations négatives entre les variables. Cette technique a été développée pour permettre l'évaluation de l'intégrale par la méthode d'échantillonnage de Monte Carlo enfin de réduire la variance.

Plusieurs auteurs ont continué à développer cette technique et ont proposé des résultats originaux dont Fishman et Huang (1983) sur la réduction de la variance avec variables antithétiques obtenue par la combinaison de la rotation et de la réflexion de l'échantillon sur un cercle unitaire.

Considérons la méthode proposée par Rubinstein (1981) qui utilise les variables u_i pour générer les x_i et $1 - u_i$ pour générer les y_i avec deux échantillons (x_i) et (y_j) où $i = 1, \dots, n$ et $j = 1, \dots, m$ distribués suivant f pour estimer I de (2.3.1).

Si $H = h \circ F^{-1}$ et $x_i = F^{-1}(u_i)$, $y_i = F^{-1}(1 - u_i)$, on a une corrélation négative entre $h(x_i)$ et $h(y_i)$ lorsque H est monotone. Mais cette condition est généralement difficile à vérifier et, de plus, cette technique ne s'applique que pour des transformations directes de variables uniformes. Geweke (1988) propose une théorie beaucoup plus originale qui consiste à créer l'inversion au niveau des x_i , en prenant $y_i = 2\mu - x_i$ lorsque f est symétrique autour de μ . Il montre sous certaines

conditions sur la fonction h , l'amélioration apportée par

$$\frac{1}{2m} \sum_{i=1}^m [h(x_i) + h(2\mu - x_i)]$$

sur

$$\frac{1}{2m} \sum_{i=1}^{2m} [h(x_i)]$$

est considérable lorsque la taille m de l'échantillon est grande. Il propose l'extension de cette approche au cas où f n'est pas symétrique, en remplaçant μ par le mode de f où la médiane de la distribution associée. De plus si f n'est pas explicite ou plus exactement si μ est inconnu, on peut approcher μ sur un premier échantillon.

Parmi les résultats fondamentaux on a parlé de la variance minimale exacte obtenue par Rubinstein et Samorodnitsky (1986) par le théorème modifié de Handscorn (1958).

On peut aussi souligner qu'il existe plusieurs applications de cette technique de variables antithétiques notamment le problème de queue traité par Mitchel (1973) et par Rubinstein (1986).

La dernière partie de ce document traite d'une nouvelle méthode pour réduire la variance avec des variables antithétiques qui consiste à estimer l'intégrale de f sur l'intervalle $(0, 1]$. L'idée fondamentale de cette méthode est de diviser cet intervalle en l sections et chaque section est subdivisée en m sections et l'objectif est de trouver le m optimal qui nous permet d'obtenir la réduction de la variance.

Par les trois résultats de simulation obtenus au troisième chapitre, on tire les conclusions suivantes :

(1) Dans l'estimation de $E(-\log(X))$, on a trouvé que la variance à l'étape 0 et à l'étape 1 est de l'ordre de n^{-3} et les trois courbes de $-\log(\text{Var})$ sont toutes décroissantes. Ce qui nous permet de prendre $m = 1$ comme valeur optimale pour réduire la variance.

(2) Pour estimer $E(X)$ par la fonction identité, les trois courbes sont parallèles donc on prendra n'importe quel m , le rapport entre deux étapes successives est égal à n^{-2} .

(3) Pour estimer $E(\exp(X))$ par la fonction identité les courbes sont croissantes et on prendra le m grand tel que $n = ml$ où n est le nombre d'observations.

(4) Pour estimer $E(\cos(tx))$ la méthode devient moins intéressante mais donne de bon résultats pour t petit ou t fixe mais n et m grands.

ANNEXE

Dans cette annexe on présente les tableaux des résultats de simulations des variances et des moyennes dans l'estimation de $E(X)$ pour le modèle exponentiel, et de $E(X)$, $E(\exp(X))$ et $E(\cos(tX))$ pour le modèle uniforme avec $t = 20$ et $n = 200$.

- Estimation de $E(X)$ pour le modèle exponentiel
- Estimation de $E(X)$ par le modèle uniforme
- Estimation de $E(e^X)$ pour le modèle uniforme

Nombres de Sections (m)	Etape 0	Etape 1	Etape 2	Etape 3
1	2,808359e-05 1,000121	2,438249e-05 0,9999332	2,438496e-05 0,9999338	2,439129e-05 0,9999339
2	2,712873e-05 0,9999618	2,810058e-05 1,0001607	2,553571e-05 0,9999991	2,545754e-05 0,9999962
4	2,654989e-05 0,9999362	3,215999e-05 1,000501	2,545408e-05 1,000028	2,436782e-05 1,000006
5	2,675149e-05 1,000038	3,581916e-05 1,000733	2,639786e-05 1,000101	2,575413e-05 1,000078
8	2,730926e-05 1,000121	4,533278e-05 1,0011780	2,709659e-05 1,0000295	2,436655e-05 0,9999586
10	2,699713e-05 1,000042	5,096515e-05 1,001612	3,459008e-05 1,000257	2,645905e-05 1,000094
20	2,508351e-05 0,9999372	7,454871e-05 1,0032290	4,463106e-05 1,0004105	2,636911e-05 0,9999387
25	2,606707e-05 0,9999613	8,716815e-05 1,0043211	4,975117e-05 1,0007135	3,627448e-05 1,0001524
40	2,53155e-05 0,9998588	1,213569e-04 1,007217	1,247872e-04 1,002467	6,196942e-05 1,000644
50	2,639776e-05 0,9999207	1,452381e-04 1,009567	1,716982e-04 1,003878	8,801164e-05 1,001191
100	2,492282e-05 0,9998932	0,0001817007 1,019417	0,0005332085 1,016312	0,0006791876 1,011085

TABLE 3.3. Annexe 1 : Tableau des variances et des moyennes pour $N = 3000$

- Estimation de $E(\cos(tX))$ pour le modèle uniforme, $t = 20$

Nombres de Sections (m)	Etape 0	Etape 1	Etape 2	Etape 3
1	2,698974e-05 1,000010	2,696673e-05 1,000040	2,707453e-05 1,000040	2,703657e-05 1,000040
4	2,698974e-05 1,000010	3,305996e-05 1,000518	2,608029e-05 1,000049	2,546416e-05 1,000032
5	2,698974e-05 1,000010	3,586239e-05 1,000695	2,661449e-05 1,000069	2,554460e-05 1,000041
8	2,698974e-05 1,000010	4,427262e-05 1,001227	2,896460e-05 1,000131	2,552147e-05 1,000050
10	2,749446e-05 1,000017	4,814801e-05 1,0015013	2,924836e-05 1,0001113	2,451638e-05 0,9999939
20	2,749446e-05 1,000017	7,310845e-05 1,003298	4,639211e-05 1,000546	2,734086e-05 1,000063
25	2,749446e-05 1,000017	8,547068e-05 1,004211	6,033392e-05 1,000870	3,297960e-05 1,000137
40	2,710683e-05 1,000011	1,204951e-04 1,007154	1,169983e-04 1,002333	6,461578e-05 1,000630

TABLE 3.4. Annexe 2 : Tableau des variances et des moyennes pour $N = 50000$

Nombres de Sections (m)	Etape 0	Etape 1	Etape 2	Etape 3
1	9,892306e-09 0,4999985	2,473077e-13 0,5	6,182691e-18 0,5	1,545672e-22 0,5
4	1,039881e-08 0,499994	2,599702e-13 0,4999625	6,499256e-18 0,4999991	1,624812e-22 0,500
5	1,026265e-08 0,500002	2,565663e-13 0,4999500	6,414157e-18 0,4999985	1,603540e-22 0,500
10	1,070328e-08 0,4999983	2,675821e-13 0,4998875	6,689552e-18 0,4999938	1,672387e-22 0,4999997
20	1,132396e-08 0,5000022	2,830989e-13 0,4997625	7,077473e-18 0,4999751	1,769368e-22 0,4999
25	1,020589e-08 0,4999972	2,551474e-13 0,4997000	6,378684e-18 0,4999610	1,594671e-22 0,4999957
40	9,724491e-09 0,4999953	2,431123e-13 0,4995125	6,077807e-18 0,4999001	1,519452e-22 0,4999800

TABLE 3.5. Annexe 3 : Tableau des variances et des moyennes ($E(X)$) pour le modèle uniforme avec $N = 1000$

Nombres de Sections (m)	Etape 0	Etape 1	Etape 2	Etape 3
1	3,411796e-08 1,718272	2,733161e-12 1,717444	2,022766e-12 1,718110	2,030135e-12 1,718248
2	3,099249e-08 1,718269	6,753063e-10 1,718261	6,734071e-10 1,718282	6,734006e-10 1,718283
4	3,521765e-08 1,718284	1,733575e-10 1,718217	1,720598e-10 1,718280	1,720567e-10 1,718282
5	3,385282e-08 1,718276	1,121810e-10 1,718196	1,111914e-10 1,718279	1,111889e-10 1,718282
8	3,295215e-08 1,718282	4,380118e-11 1,718132	4,244908e-11 1,718275	4,244697e-11 1,718282
10	3,396605e-08 1,718280	2,756516e-11 1,718088	2,670024e-11 1,718271	2,669207e-11 1,718281
20	3,150728e-08 1,718276	8,361266e-12 1,717874	7,718320e-12 1,718239	7,722800e-12 1,718277
25	3,351872e-08 1,718278	5,690649e-12 1,717766	4,779766e-12 1,718215	4,780094e-12 1,718273
40	3,159933e-08 1,718291	2,573192e-12 1,717444	1,990183e-12 1,718110	1,990389e-12 1,718247
50	3,071925e-08 1,718282	1,901965e-12 1,717230	1,326000e-12 1,718013	1,334954e-12 1,718215
100	3,406762e-08 1,718275	1,215152e-12 1,716156	4,250913e-13 1,717208	4,466079e-13 1,717745

TABLE 3.6. Annexe 4 : Tableau des variances et des moyennes de $(E(\exp(X)))$ par le modèle uniforme avec $N = 1000$

Nombres de Sections (m)	Etape 0	Etape 1	Etape 2	Etape 3
1	2,095013e-06 0,04561327	2,013716e-06 0,04562024	2,013945e-06 0,04562063	2,013950e-06 0,04562063
2	2,109782e-06 0,045605	1,808650e-06 0,04568123	1,806865e-06 0,04567467	1,806877e-06 0,04567457
4	2,028935e-06 0,04564531	1,971921e-06 0,04561044	1,976948e-06 0,04558923	1,976998e-06 0,04558868
5	2,022686e-06 0,04566555	1,442144e-06 0,04568028	1,444088e-06 0,04565414	1,444100e-06 0,04565326
8	1,972712e-06 0,04559531	9,801088e-07 0,04572146	9,783677e-07 0,04567229	9,782928e-07 0,04566995
10	2,10688e-06 0,04565702	6,745445e-07 0,04573984	6,795794e-07 0,04567692	6,796415e-07 0,04567339
20	2,058936e-06 0,04567032	1,775460e-07 0,04575878	1,791444e-07 0,04563292	1,789865e-07 0,04561962
25	1,967160e-06 0,04562044	1,251493e-07 0,04582289	1,269363e-07 0,04566815	1,269011e-07 0,04564811
40	2,009222e-06 0,04560486	5,381374e-08 0,04593410	5,643721e-08 0,04570343	5,673891e-08 0,04565610
50	2,102940e-06 0,04564217	3,209352e-08 0,04600342	3,397609e-08 0,04573417	3,420755e-08 0,04566502
100	2,043632e-06 0,04566561	7,643561e-09 0,04637411	1,017092e-08 0,04601797	1,082872e-08 0,04583455

TABLE 3.7. Annexe 5 : Tableau des variances et des moyennes de $(E(\cos(tX)))$ pour le modèle uniforme avec $N = 1000$ et $t = 20$

PROGRAMME PLUS

Initialisation

```
n<-200 # Nombre d'observations
ℓ<-200 # Nombre de Sections
N<-1000 # Nombre de Simulations
m<-1 # Nombre de sous-intervalles /chaque section
h<-3 # Nombre d'étapes
t<-20
Yo<-c()
z<-c()
ℓ4<-c()
Zo<-NULL
Lo<-NULL
ro<-NULL
r<-NULL
d<-c()
ra<-c()
ra0<-NULL
d1<-matrix(NA,nrow=m,ncol=ℓ)
v<-c()
u<-matrix(NA,nrow=m,ncol=ℓ)
R<-matrix(NA,nrow=m,ncol=ℓ)
a<-matrix(NA,nrow=m,ncol=ℓ)
Xo<-matrix(NA,nrow=m,ncol=ℓ)
```

```

p<-c()
x<-matrix(NA,nrow=m,ncol=l)
y<-matrix(NA,nrow=h,ncol=N)

```

Simulations

```

*****

```

```

for ( j in 1 :N )

```

Etape 0

```

*****

```

```

for ( r in 1 :n){
    d[r]<-(r-1)/n
}
a <-matrix(d,nrow=m,byrow=T)
for ( s in 1 :m){
    for ( i in 1 :l){
        u[s,i]<-runif(1,0,1)
        R[s,i]<- u[s,i]/(m*l)
        Xo[s,i]<-a[s,i]+R[s,i]
    }
}
Yo[j]<-mean(-log(Xo))

```

Etape $k \geq 1$

```

*****

```

```

for ( k in 1 :h){

```

Génération d'une permutation

```

*****

```

```

for ( s in 1 :m){
    for ( i in 1 :l){
        v[i]<-runif(1,0,1)
    }
}

```

```

    p<-order(v)
  for ( i in 1 : $\ell$ ){
    d1[s,i]<-(p[i]-1)/(m*( $\ell^{(k+1)}$ ))
    R[s,i]<-R[s,i]/(m* $\ell$ )
    a[s,i]<-a[s,i]+d1[s,i]
    x[s,i]<-a[s,i]+R[s,i]
  }
}
y[k,j]<-mean(-log(x))
}
}

for ( k in 1 :h){
  Moyenne à chaque étape de  $k \geq 1$ 
  *****

  z[k]<-mean(y[k,])

  La variance à chaque étape de  $k \geq 1$ 
  *****

   $\ell_4[k]$ <-var(y[k,])
}

Moyenne à l' étape 0
*****

Zo<-mean(Yo)

La variance à l'étape 0
*****

Lo<-var(Yo)

```

BIBLIOGRAPHIE

- [1] $\mathcal{A}\mathcal{M}\mathcal{S}\mathcal{I}\mathcal{A}\mathcal{T}\mathcal{E}\mathcal{X}$ VERSION 1.1 USER'S GUIDE, Amer. Math. Soc., Providence, R. I., 1991.
- [2] Casella, G. & Berger, L.R., (2002), *Statistical Inference*, Second edition, Duxbury, Pacific Grove, CA.
- [3] Cheng, R.C.H., (1982), The use of antithetic variates in computer simulations, *J. Oper. Res. Soc.*, **33**, 229-237.
- [4] Devroye, L., (1986), *Nonuniform random variate generation*, Springer-Verlag, New York.
- [5] Fishman, G., (1996), *Monte Carlo : Concepts, algorithms, and applications.*, Springer Series in Operations Research, Springer-Verlag, New York.
- [6] Fishman, G. & Huang, B.S. D., (1983), Antithetic variates revisited, *Comm. ACM* , 26, **11**, 964-971.
- [7] Gal,S., Rubinstein, R.Y, & Ziv A., (1984), On the optimality and efficiency of common random numbers, *Math. Comput. Simulation* 26, **6**, 502–512.
- [8] Gentle, J.E., (1998), *Random number generation and Monte carlo methods, statistics computing*, Springer, New York.

- [9] Goossens, M., Mittelbach, F. & Samarin, A., (1994), *The L^AT_EX companion*, Addison-Wesley, Boston.
- [10] Hammersley, J.M. & Handscomb, D.C. (1964), *Monte Carlo methods*, Wiley, New York.
- [11] Hammersley, J.M. & Mauldon, J.G. (1956), General principles of antithetic variates, *Proc. Cambridge Philos. Soc.*, **52**, 476-481.
- [12] Hammersley, J.M. & Morton, K.W. (1956), A new Monte Carlo technique : antithetic variates, *Proc. Cambridge Philos. Soc.*, **52**, 449-475.
- [13] Kinderman, A.J. & Monahan, F.J.(1987), Computer generation of random variables using the ratio of uniform deviates, *ACM Trans. Math. Software*, **3**, 3, 257-260.
- [14] Law A.M. & Kelton, W.D. (2000), *Simulation Modeling and Analysis : Third Edition*, McGraw-Hill series, New York.
- [15] Lehman, R.S. (1977), *Computer Simulation and Modeling : An Introduction*, John Wiley Serie.
- [16] Leydold, J. (2000), Automatic Sampling with the Ratio-of-uniforms Method, *ACM Transactions on Mathematical Software* **26**,(1), pp. 78-98.
- [17] Madras, N. (2002), *Lectures on Monte Carlo methods*, American Mathematical Society, Fields Institute Monographs.
- [18] Mitchel, B.(1973), Variance Reduction by Antithetic Variate in GI/G/1 Queueing Simulation, *Oper.Res.*, 21, 988-997.

- [19] Morton, K.W. (1957), A generalisation of the antithetic variate technique for evaluating integrals, *Journal Mathematical physics* **36**, 289-293.
- [20] Robert, C., (1986) *Méthodes de Monte Carlo par chaines de Markov*, Edition Economica, Paris.
- [21] Rice, J.A. (1995), *Mathematical Statistics and data analysis*, Duxbury Press, Belmon, California, Second edition.
- [22] Rubinstein, Y.R. & Samorodnitsky G. (1987), A modified version of Handscomb's antithetic variates theorem. *SIAM J. Stat. Sci. Comp. Simul.*, **8**, 82-98.
- [23] Rubinstein Y.R.(1981), *Simulation and Monte Carlo methods*, Wiley Series, New-York.
- [24] Rubinstein Y.R.(1986), *Monte Carlo optimization, Simulation and Sensitivity of queueing Networks*, Wiley Series, New-York.
- [25] Rubinstein, Y.R., Samorodnitsky, G. & Shaked, M. (1985), Antithetic variates, multivariate dependence and simulation of stochastic systems, *Management Sci.*, **31**, No 1, 66-77.
- [26] Snijders, T.A.B., (1984), Antithetic variates for Monte Carlo estimation of Probabilities, *Statistic Neerlandica*, **38**, No 2, 1-19.
- [27] Ștefănescu, Ș. & Văduva, I., On computer generation of random vectors by transformations of uniformly distributed vectors. of random vectors by transformations of uniformly distributed vectors, *Computing*, **39**, 141-153.
- [28] Uhinov, S.A. et Voytishek, A.V.(2000), Usage of the importance sample in Monte Carlo methods, *Monte Carlo methods applications*, **6**, No 3, 341-348.

- [29] Wilson, J. R. (1979), Proof of the antithetic-variates theorem for unbounded functions, *Math. Proc. Cambridge Philos. Soc.*, **86**, 3, 477-479.